



BUSINESS IN TIME
Atlantic Schools of Business
StFX UNIVERSITY | 2013

43rd Annual
Atlantic Schools of Business Conference
2013 Conference Proceedings

St. Francis Xavier University
Antigonish, Nova Scotia
September 27 - 29


ST. FRANCIS XAVIER
UNIVERSITY



ASB 2013
Antigonish, Nova Scotia

Kirk A. Collins
Schwartz School of Business
St. Francis Xavier University

Joshua Sim (graduate student)
Department of Economics
Queen's University

TAXATION THROUGH TIME: AN INTERTEMPORAL ASSESSMENT OF THE TAXATION OF HUMAN CAPITAL IN ATLANTIC CANADA*

We examine the (dis)incentive effects created by the respective tax systems to invest in human capital in Atlantic Canada and compare this to a select group of provinces from the rest of Canada. While findings show a steady decline in effective tax rates through the years, thereby creating an incentive effect to invest in post-secondary education, disproportionately higher rate gap differentials in the Atlantic Provinces, on average, combined with negative comparative statics reveal a somewhat different undertone. The counterproductive nature of the competing policies effectively nullifies any status quo argument for education or tax policy in the Atlantic Provinces, when compared to their brethren. The graduate retention rebate provides some solace to the narrative in helping to alleviate early tax burdens and equalize returns, but is difficult to claim effectiveness in assuaging any monetary windfall associated with migration.

Introduction

The goal of this paper is to assess the (dis)incentive effects created by the provincial tax systems for human capital investment in Atlantic Canada. Our test subjects are university graduates, but we note that such methodology could be applied to any level or subset of the population that possess human capital deemed scarce and productive, in the economic sense. We do not purport to suggest that university graduates are more important in the economic well being of an economy than, say, skilled trades or college graduates, e.g., and encourage one to think of our case as a subset of a larger sample of possibilities. Our presumption, like our goal, is simple; namely, an empirical investigation of incentivizing the tax system in Atlantic Canada remains an unexplored possibility for strategic policy implementation to retain and recruit scarce resources in the market for human capital. We use the impact on effective tax rates (ETRs) and rates of return as our policy litmus test and explore by empirical and simulation analysis the unilateral exploitation of a select number of provincial tax systems. Some may claim that this work puts the

* Collins would like to acknowledge the support of the CRS-SSHRC institutional research grant at St. Francis Xavier University in helping to complete this research.

proverbial cart before the horse, as it were, since to attract workers a province needs the requisite infrastructure and such may not be currently present at the appropriate level(s) to warrant an increase in the supply of skilled labour. To that end, we suggest that a more accurate assessment (or, at least, a plausible alternative) might be akin to a “chicken or the egg” argument; namely, which comes first? We leave this debate up to the interested reader for now, and focus on establishing the empirics of the analysis in an effort to add the requisite fodder for a more informed discussion.

While previous research has examined the disincentive effects created by the tax system for, e.g., university and college graduates (see, for instance, Burbidge et al., 2012), the current study is differentiated in a number of key ways. First, the level of detail contained within Kevin Milligan’s 2012 *Canadian Tax and Credit Simulator* (CTaCS) allows us to model investment decisions for the provinces in question for an extended period of time (1962-2012), which other studies have simply not done. The focus on creating a comparative advantage for human capital investment in Atlantic Canada also distinguishes this work from its predecessors, since much of the previous research is confined to Ontario as the provincial counterfactual (e.g. Collins and Davies, 2004, 2005a/b). Research that has considered provincial implications for human capital taxation, such as Collins (2008) and Burbidge et al. (2012), have done so by means of an aggregated snapshot of one or two years. The use of the CTaCS database ensures that all of the idiosyncratic deductions are accounted for in determining the full treatment that the tax system implies for the respective provinces. For instance, we account for items such as the graduate tax credit and graduate retention rebates, whereas previous studies stopped short of examining these aspects. Finally, as means to illustrate the impact of the graduated rate structure on human capital investment, the CTaCS database affords a number of simulation opportunities to examine the implications of having an individual graduate earn at various quantiles over their lifecycle. To provide a degree of semblance with previous studies and add further fodder for discussion we also use earnings data from the 2006 Census for our empirical analysis.

Findings show that tax systems in the Atlantic Provinces tend to create additional disincentive effects, when compared to non-Atlantic; Nova Scotia and PEI have a particularly ominous outlook for taxation of human capital over the last 50 years in that their treatment remains comparatively poor, when examined in relation to other provinces in our study. Of course, this does not imply that the tax treatment of human capital has not become more favorable, when compared to, say, its treatment in 1962; indeed it has. What it is meant to imply is that when compared with what other provinces have done in terms of tax policy, these two, in particular, consistently find themselves at the highest end of the effective tax rate spectrum, which is a rather unenviable position.

In addition, we find an interesting dichotomy in our bilateral comparison of Nova Scotia and Ontario graduates. In particular, items such as the graduate tax credit and the more recent graduate retention rebate go some way to reversing the disincentive effect associated with the higher rate structures in Nova Scotia, returning both males and females to an after-tax return comparable to their Ontario counterparts. Of course, the design of the graduate retention rebate was not to reimburse individuals to a point of return neutrality, if you will. It was designed as an

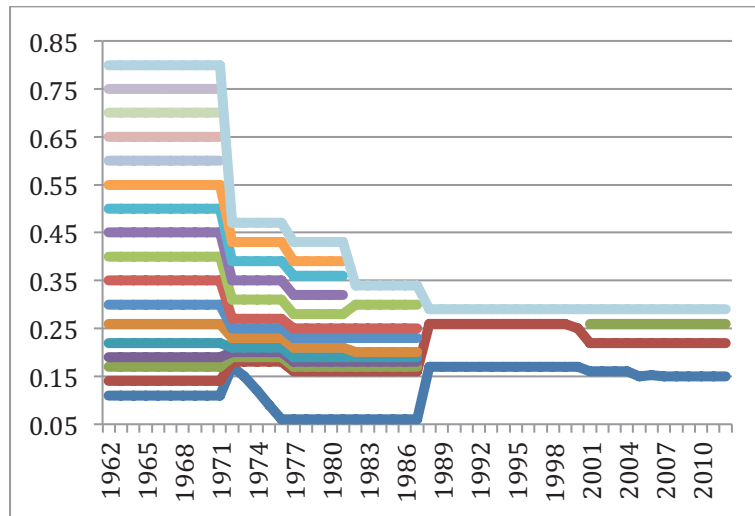
incentive to remain in or to potentially relocate to the respective province. In this regard, the current policy appears somewhat unsuccessful in its mandate.

The rest of the paper is organized as follows. In the next section, we provide some motivation and a primer on relevant tax implications. We follow this with a discussion of the methodology used to assess the tax (dis)incentive effects for human capital investment. Our results follow and, finally, we conclude with some remarks and avenues for future research.

Motivation and Primer

Taxation has changed remarkably from the time of Carter (Figure 1). Provincial governments have largely followed the lead of the federal government, by first applying a tax on tax approach to collecting revenue before migrating to the tax on net income approach used today (see Appendix for a detailed look at the Atlantic Provinces). Deductions have increased and surtaxes have, for the most part, decreased or been eliminated.

Figure 1
Federal Tax Rates, 1962-2012



Source: Kevin Milligan's 2012 CTaCS database

During this time the corporate income tax has also undergone its own transformation, starting out at greater than 2.5 times where it currently sits (15% Federal rate). What is more is that the corporate rate has experienced a far greater reduction over the last fifty years, than the personal income tax. The result is the appearance of an increasing rate gap, since the early nineties; the corporate rate has fallen by approximately 46%, while the top personal rate has remained unchanged (Table 1).

Table 1
Federal Rate Gap, 1962-2012

	Top PIT Rate	CIT	Rate Gap
1962	80	41	39
1972	80	40	40
1982	43	36	7
1992	29	28	1
2002	29	25	4
2012	29	15	14

Sources: Authors' calculations, CTaCS database and Collins and Edgar (forthcoming)

Within the provincial landscape there are issues of inequity, as expected, but the rate gap still persists for all but one province; that being Newfoundland. It is also worth pointing out that only Nova Scotia finds itself in a position of having a rate gap anywhere close to the federal one, on any measurement (see, the gap for small businesses, Table 2). The issue is, unfortunately, systematic of a larger problem. As authors have pointed out, for many countries the combination of imperfectly mobile physical capital and international tax competition has resulted in diverging corporate and personal rates (See, Head and Krever, 2009). Vertical equity and social policy being espoused as reason why the personal rates remain so high and create the all too familiar incentive compatibility problem for corporations where income shifting and tax avoidance, through corporate retentions, are concerned. The flip side to this inequity is that human capital is also affected by the resulting policy dynamic.

Table 2
Federal and Provincial Rate Gap, 2013

	Federal	BC	AB	ON	NB	NS	PE	NL
Corporate	15	11	10	11.5	11	16	16	14
M&P	15	10	10	10	10	16	16	5
Small Bus.	11	2.5	3	4.5	4.5	3.5	4.5	4
Top PTR	29	14.7	10	13.16	16.07	21	16.7	13.3
Corp Rate Gap	14	3.7	0	1.7	5.1	5	0.7	-0.7
M&P Rate Gap	14	4.7	0	3.2	6.1	5	0.7	8.3
Small Bus Rate Gap	18	12.2	7	8.7	11.6	17.5	12.2	9.3

Source: Authors' calculations, CTaCS database, and
<http://www.novascotia.ca/finance/en/home/taxation/othercomparativetaxrates.aspx>

Where human capital is concerned, this classic treatment of corporate income tax on inbound investment, at rates lower than the highest personal income tax rate, creates a hold-up problem of sorts, where the government must collect the necessary revenue from alternative sources to fund social policy. The notion is to collect it by means of tax on personal income. The problem is that human capital is, to a large extent, mobile. And it is this mobility that has policymakers searching for viable alternatives to stem the flow of their best and brightest.

Using the tax system as a means to capitalize on the externalities wrought by an economy that is high in human capital is something that has recently started to take shape. For instance, many provinces now provide a graduate retention supplement or rebate for those relocating or holding permanent residency after graduating from an approved institution. We perform an assessment of one such program below (for Nova Scotia), but suffice it to say, that the piecemeal approach with which such enactment has occurred, while creating (perhaps) a comparative advantage for some provinces, falls dramatically short in making any kind of substantial impact on the decisions of graduates to relocate. The inertia from relocating would have to be sufficiently weak (be that through, say, increased opportunities, money, etc.), for the current format of the policy to even enter the decision-making of recent graduates. Anecdotally, the policy appears more akin to a graduation gift, of sorts, for remaining where one already intended, since it does little in the way to level any so-called playing field, as we show below.

Methodology

Our examination of the disincentive effects of taxation on human capital investment uses the effective tax rate methodology of Collins and Davies (2004). Briefly, the methodology is predicated on the notion that the proportional reduction in after-tax internal rates of return, which manifest as a result of the inherent tax structure, will permit a snapshot of how much influence the latter exerts on lifetime earnings. The greater the effective tax rate, the larger the disincentive effect to invest in that particular locale. Results are typically treated as a steady state due to limited empirical observations. While we maintain this assumption in order to provide some semblance with previous empirical studies, we also extend the analysis to take into account associated year effects.

Our earnings profiles are generated from the quantile regression of Burbidge et al. (2012), who use quartic polynomials in age, and the interested reader is directed to their study for a complete assessment. We limited ourselves to median earners for this study, but note that any quantile of the distribution could theoretically be assessed. To ensure consistency across the provinces in terms of tax treatment, the counterfactual earnings profile of the median university graduate is used. We recognize that this may bias the results insofar as earnings capacity differs, thereby permitting a bracket creep, of sorts. Our methodology is designed to confine our results to tax implications in a comparative statics environment, since our focus is on the idiosyncratic nature of the respective tax systems and not the magnitude of the earnings differential, should one exist.

For the empirical results, data on tuition and additional expenses are taken from Statistics Canada.¹

Results

Canadian Tax and Credit Simulator (CTaCS) Analysis

Using Kevin Milligan's 2012 CTaCS database, we perform a comparative statics exercise on the provinces listed in Table 3. The experiment is designed as a worst-case scenario of disincentive effects associated with the respective tax systems, since it places each graduate at the top rate upon graduation and keeps them there over their working lifetime. We start by ensuring the same before tax rate of return in each of the years listed, and then run a simulation program (in SAS) that applies all tax and education credits available to determine after-tax rates of return. The resulting ETR values are listed below.

Table 3
ETRs by Province, CTaCS Simulation, 1962-2012

	BC	AB	ON	NB	NS	PE	NL
1962	27.46%	26.25%	27.54%	28.86%	28.52%	28.18%	29.10%
1972	23.15%	21.51%	23.59%	24.21%	24.71%	24.24%	24.78%
1982	19.54%	18.61%	20.29%	20.77%	21.75%	21.41%	21.18%
1992	15.89%	15.67%	16.79%	17.49%	18.98%	18.59%	17.76%
2002	13.64%	13.57%	15.04%	15.50%	17.27%	17.07%	15.39%
2012	11.96%	11.80%	13.58%	13.88%	15.74%	15.80%	13.73%

Source: Authors' calculations

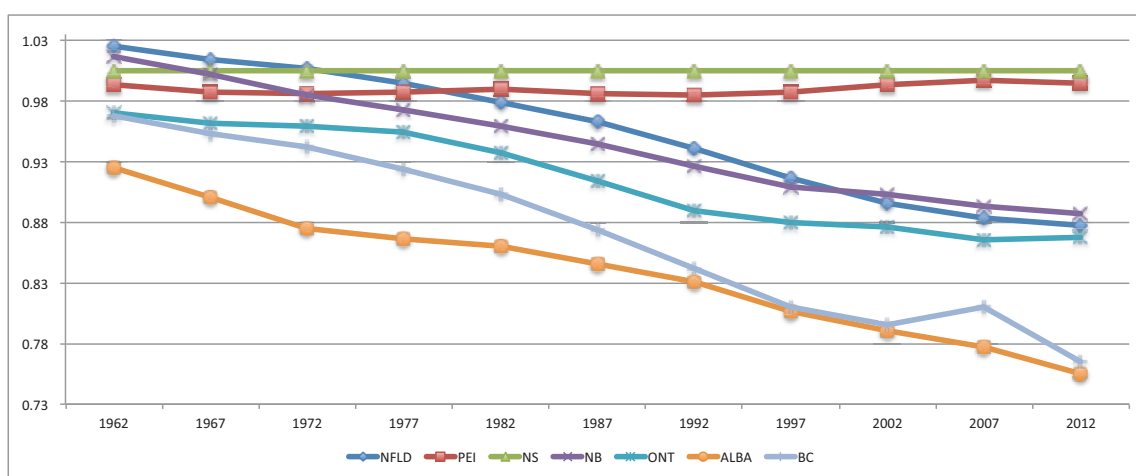
As is evident from Table 3, the situation of the top earners has progressively gotten better over the years. This outcome is largely due to a reduction in the progressivity of the personal income tax system, as opposed to by design, where education policy is concerned; the resulting benefit, in other words, is simply a positive externality of a tax system undergoing refinement. Fortunately, many deductions specifically earmarked for education start to appear in latter years and this, in turn, helps to alleviate the burden otherwise imposed by personal income tax.

PEI and Nova Scotia are amongst those that create the largest disincentive effect, while British Columbia and Alberta find themselves at the opposite end. Upon first blush it may appear that the rates are somewhat similar in magnitude, but consider the analysis in Figure 2. It continues with the steady state assumption of Table 3, but provides comparison via a benchmark (Nova Scotia) to afford a different look at the relative numbers. The other provinces are listed relative to their

¹ See Tables, <http://www.statcan.gc.ca/daily-quotidien/060901/t060901a-eng.htm> [Accessed, July 17, 2013]. Alternatively, Table 12, Burbidge et al. (2012) contains a list of all relevant 2006 cost and tax parameters.

ETR valuation to the benchmark. So, for example, a graduate in Alberta in 2012, assuming a steady state, would find him/herself benefiting from a 25% reduction in taxation of returns to post-secondary education, when compared to Nova Scotia. As the figure clearly indicates, the situation for graduates, while favorable when viewed over the last fifty years, is somewhat less so if one is located in the Atlantic Provinces, especially Nova Scotia and PEI.

Figure 2
Proportional ETR Comparative Statics Analysis, CTaCS Simulation, 1962-2012
 (NS = 1.00)



Source: Authors' calculations

Empirical Analysis

The empirical analysis uses the tax system and associated education deductions in 2006 to derive after-tax earnings profiles of median male and female university graduates. For this paper, we explore the case of Ontario and Nova Scotia and assess the effects therein, but note that any provincial comparison is viable. In the case of Ontario this implies a three bracket system (6.05%, 9.15%, and 11.16%), compared to Nova Scotia's four (8.79%, 14.95%, 16.67% and 17.50%), with associated income thresholds of \$34,758 and \$69,517, and \$29,590, \$59,180 and \$93,000, respectively. Basic personal amounts are \$8,377 for Ontario, credited at 6.05%, with a textbook credit of \$451. For Nova Scotia, comparable amounts are \$7,231, 8.79% and \$200. Direct costs are \$5,160 for tuition and \$729 for additional fees in Ontario. In Nova Scotia, these costs are \$6,571 and \$572, respectively. Other expenses are assumed to be \$1,000, for both provinces.

The first two columns of Tables 4 and 5 represent our base case, for males and females, respectively, for Ontario and Nova Scotia and are computed via the numbers reported above. The remaining columns in our Nova Scotia case, if you will, are dealing with the added education benefit associated with the (previously in place) graduate tax credit (GTC) and (current) graduate

retention rebate (GRR). For the base case, these deductions were withheld in computing the impact of the provincial tax system in an effort to make it more analogous to the Ontario treatment. The subsequent columns add these deductions back to examine the impact on incentive effects. We present sensitivity analysis around the GTC by eliminating the tax burden completely in the graduation year (GTC-All Tax). The experiment is designed, in part, to show the idiosyncratic nature of policy changes on males and females; note that the difference means very little to females, but has a larger impact on males. More on this below.

Table 4
Median University Graduate, Male, 2006

	Ontario	Nova Scotia			
		Base Case	GTC	GTC- All Tax	GRR (6yr)
Before-Tax IRR	17.9	17.1	17.1	17.1	17.1
After-Tax IRR	16.7	15.5	15.7	15.9	16.6
ETR	6.9	9.4	8.5	7.1	3.0

Source: Authors' calculations

The results show that the GRR is far better at reducing the disincentive effects associated with the tax system and investing in human capital, than the previous GTC. The findings also highlight idiosyncrasies with respect to the inequitable treatment of the tax system that are always present. In this case, though, the numbers are particularly interesting, especially insofar as the current (non-refundable) graduate retention rebate is concerned.

Table 5
Median University Graduate, Female, 2006

	Ontario	Nova Scotia			
		Base Case	GTC	GTC- All Tax	GRR (6yr)
Before-Tax IRR	22.0	20.8	20.8	20.8	20.8
After-Tax IRR	21.1	19.6	19.9	19.9	21.0
ETR	4.2	5.5	4.4	4.3	-0.9

Source: Authors' calculations

That is, despite differentiations in earnings (and the associated premiums) that result in higher rates of returns for females over males, a GRR of six year returns both of these median graduates to a(n) (approximately) similar after-tax rate of return, when compared to their Ontario counterparts. This is a rather magnanimous result and one that would seem to support a narrative of pro-retention (not necessarily, recruitment) policy. Unfortunately, when migration implications are modeled, this result falls apart somewhat (see Table 6).

Table 6 shows that the impact for those in Nova Scotia considering a move to Ontario are exacerbated due to the less progressive system; in this case, graduates in Nova Scotia are effectively being subsidized due to the differential tax treatment, to migrate. Combine this with any realized increase in compensation over the median graduate as a result of being located in Ontario and the magnitude of the difference increases further. Note that the reverse holds, as well, for those in Ontario considering moving to Nova Scotia after graduation.

Table 6
Median University Graduates, Male, Migration Results, 2006

	Ontario to Nova Scotia				Nova Scotia to Ontario
	w/o GTC/GRR	w/ GTC	w/GRR	from NS	
Before-Tax IRR	17.9	17.9	17.9	17.9	17.1
After-Tax IRR	14.9	15.1	16.0	17.0	17.2
ETR	16.6	15.9	10.6	4.8	-0.2

Source: Authors' calculations

The current policy also creates an incentive effect for graduates, who ultimately desire to remain in Nova Scotia, to seek education *outside of the province*, since in 2006 Nova Scotia had the highest tuition of any Canadian province (see, e.g., Table 12, Burbidge et al, 2012). Going to school out-of-province would effectively increase the rate of return due to the realized decrease in outflows associated with obtaining the education. Furthermore, once opportunity costs of migration are taken into consideration, Table 6 shows that the retention and recruitment aspects of the GRR policy appear moot, given the resulting lifecycle windfall from leaving. As such, the policy in place seems more akin to rewarding those who had no intention of going anywhere, for staying.

Taken as a whole, there appears to be an incentive compatibility problem brewing in education and tax policy in Nova Scotia, particularly, if as is often espoused, such policies are designed with retention and recruitment in mind (see, e.g., CCH's Provincial Tax News, no. 5, October 2011). Consider the fact that the cost of the GRR in its first year (2009) was approximately \$3.9m, compared to the cost of the GTC of \$1.4m (NS Dept of Finance, Overview of the Nova Scotia Tax System, April 2011), then the implications of forgoing additional revenues take on a rather sinister feel, assuming no subsequent inflow from out of province has resulted; when considered in conjunction with the funding cuts of post secondary institutions in the province, the incumbent policies appear even more stark.

The above notwithstanding, the creation of the GRR is without a doubt one of the more innovative features in reducing the disincentive effect associated with the tax system in terms of human capital investment. The impact on post-graduate earnings is the key to its success and this is the most important thing to keep in mind when any revamping or further changes are to be considered. It is the impact of the tax system on the earnings lifecycle after graduation that matters; that is, after the human capital has been acquired, and not while the individual is in school. While it is true that education and textbook amounts lower ETRs, what has a far greater

impact is reducing the tax burden after graduation. Provinces would be better served at reducing tax-based incentives of individuals while they are in school and focus more on alleviating the burden when they enter the marketplace. In this respect, we offer the following policy experiments in Table 7 for additional reference.

Table 7
Median University Graduates, Male, Policy Simulation, 2006

	Base Case	Deregulated Tuition (\$10,000/yr)		
		No GRR	GRR (6yr)	GRR (14yr)
Before-Tax IRR	17.1	15.4	15.4	15.4
After-Tax IRR	15.5	13.9	14.9	15.5
ETR	9.4	9.8	3.0	-1.0

Source: Authors' calculations

Table 7 effectively deregulates tuition to the tune of \$10,000 per year in 2006. Not surprisingly there is a noticeable impact on rates of return, but note the marginal impact on ETRs when no graduate retention rebate (GRR) is being offered. Again, the illustration is meant to show that anything done to students while they are in university pales in comparison to how they are treated post-graduation. Columns 3 and 4 of Table 7 attest to this latter fact. Despite the approximately 50% increase in tuition in 2006 terms, a graduate retention rebate of 14 years would effectively nullify the front loaded expense by returning the individual to the same after-tax return of 15.5%. In other words, a total increase in tuition of approximately \$7,000 realized over four years would effectively be eliminated by a tax reduction of \$35,000 spread out over 14 years; ignoring time value of money, which would effectively lower the foregone amount.

Assuming part-time students represent a third of full-time, then there were roughly 37,300 FTE university students in Nova Scotia in 2006.² With each of these students paying an extra \$3,500 that equates to roughly \$130.6m. If, and it is a big if, the GRR for 2009 is indicative of the approximate yearly cost to run the program then the outflow would be roughly \$56m for the 14 years, ignoring time value of money and any growth in program expenses. In this case universities could essentially pay for the added costs, if you will, associated with the lengthening of the GRR program by means of their increased revenue and still have a substantial amount left over to fund their institution's mandate. Note that if there were a substantial increase in uptake beyond the revenue received upfront, it would necessarily imply an influx of human capital to the province that would translate into additional revenue from personal income taxes, sales taxes, etc., so the outcome might still warrant consideration of such policy changes.

² See, <http://www.statcan.gc.ca/daily-quotidien/090311/t090311a3-eng.htm> [accessed July 17, 2013].

Concluding Remarks

We have attempted to shed light on the relative (dis)incentive effects created by the provincial tax systems for the Atlantic Provinces. While recent measures, such as the graduate retention rebate, go some way in mitigating years of relatively high disincentive effects in relation to provinces outside of the region, they still do not do enough to impact migration incentives to any real extent. The current system in Nova Scotia also creates additional returns for graduates to receive their education outside of the province and return home, as opposed to promoting in-province education, which is rather vexing. Whether or not other this trend persists for the other provinces in Atlantic Canada is a recommendation for further research; likewise for other educational providers besides universities.

Furthermore, while increases in tuition are often espoused as unfairly credit constraining (or handicapping) future graduates, there is an argument in favor of deregulating tuition in conjunction with mitigating tax burdens upon graduation for a realized benefit. Of course, we recognize that our return-based approach does not fully encompass the credit issue. Nor does it model the subsequent changes in (opportunity and real) costs in tax revenue and expenditures that would manifest. As such, we would once again encourage more research in this important area to determine the overall impact.

Finally, while our counterfactual was to make median graduates as well off in an after-tax sense, other measures, outside of monetary ones, are also relevant to recent graduates in terms of their migration decisions (such as proximity to family and friends, job opportunities, etc.). Whether or not such characteristics override monetary ones would require a rather extensive framework to measure appropriately, but nonetheless is an interesting avenue for future research.

References

Boothby, D., and T. Drewes (2006) "Post-secondary education in Canada: returns to university, college and trades education," *Canadian Public Policy* 32, 1–22

Burbidge, J., K.A. Collins, J.B. Davies, and L. Magee (2012) "Effective Tax and Subsidy Rates on Human Capital in Canada," *Canadian Journal of Economics*, 45(1), 189-219.

Collins, K.A. (2008) "The 'Taxing' Issue of Interprovincial and Cross-Border Migration," *Canadian Public Policy* 34(4), 481-500.

Collins, K.A. and J.B. Davies (2004), "Measuring Effective Tax Rates on Human Capital: Methodology and an Application to Canada", in Peter Birch Sorensen (ed.), *Measuring the Tax Burden on Capital and Labor*, MIT Press, Cambridge, Mass. pp. 171-211.

Collins, K.A. and J.B. Davies (2005a), "Carrots and Sticks: The Effect of Recent Spending and Tax Changes on the Incentive to Attend University," C.D. Howe Institute, *Commentary*, 220, October, 22 pp.

Collins, K.A. and J.B. Davies (2005b), "Recent Changes in Effective Tax Rates on PSE Level Human Capital in Canada," in (eds.) Charles Beach, Robin Boadway, and Marvin McInnis, *Higher Education in Canada*, Queen's University Press, Kingston, Ontario, pp.417-451.

Collins, K.A. and T. Edgar (Forthcoming), "The Different Roles of Corporate Income Tax: Assessing the Current Canadian System against the Carter Ideal and Other Reform Alternatives," *Tax Reform: Fifty Years in the Making*, Toronto: Carswell.

Day, K. and S.L. Winer (2006) "Internal Migration and Public Policy in Canada: An Empirical Study," *International Tax and Public Policy* 13, 535-564.

Head, J.G and R. Krever (2009) "Australian Business Income Tax Reform in Retrospect: An Analytical Perspective" in Chris Evans and Richard Krever, eds, *Australian Business Tax Reform in Retrospect and Prospect* (Sydney: Thomson, Reuters) 17.

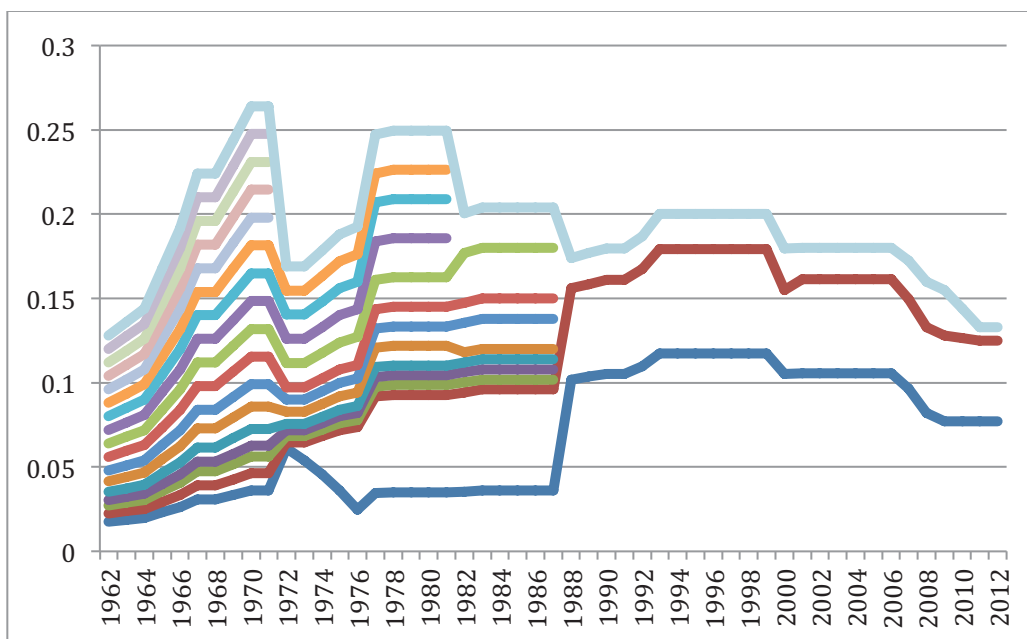
Milligan, K. (2012), *Canadian Tax and Credit Simulator*. Database, software and documentation, Version 2012-1.

Vaillancourt, F., and S. Bourdeau-Primeau (2002) "Returns to university education in Canada, 1990 and 1995," in *Renovating the Ivory Tower: Canadian Universities and the Knowledge Economy*, Policy Study 37, ed. David E.W. Laidler (Toronto: C.D. Howe Institute)

Appendix

The tables below represent the evolution of tax rates within the respective provinces for the years 1962-2012. As conveyed above, the rates mimic the federal ones, if you will, up until 2000, when there was a switch from the tax on tax approach to a tax on net income one. The Tables are provided for information and derived from Kevin Milligan's 2012 CTaCS database.

Table 1A
Newfoundland Tax Rates, 1962-2012



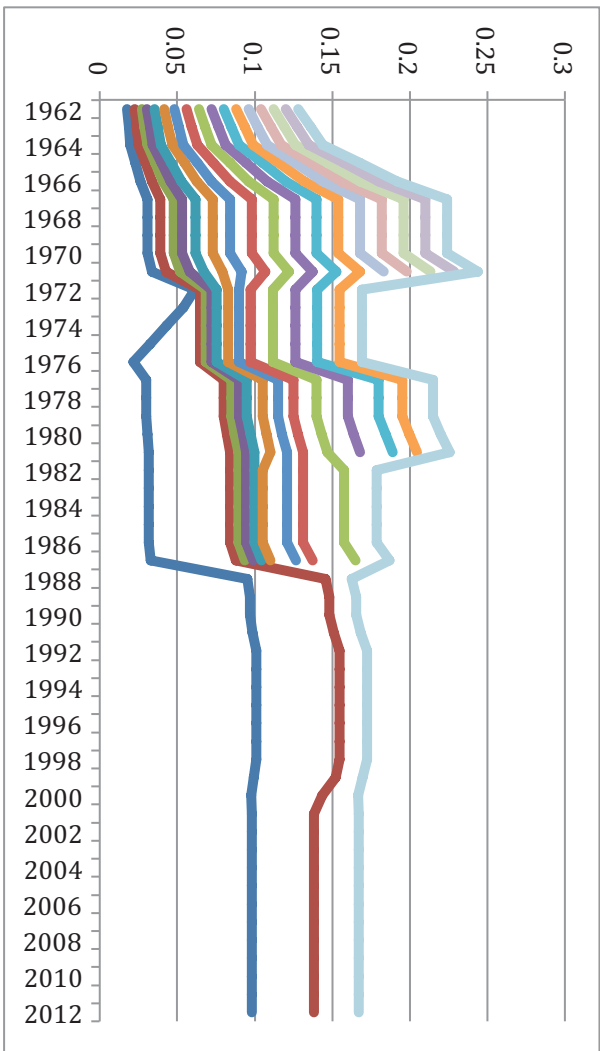


Table 2A
PEI Tax Rates, 1962-2012

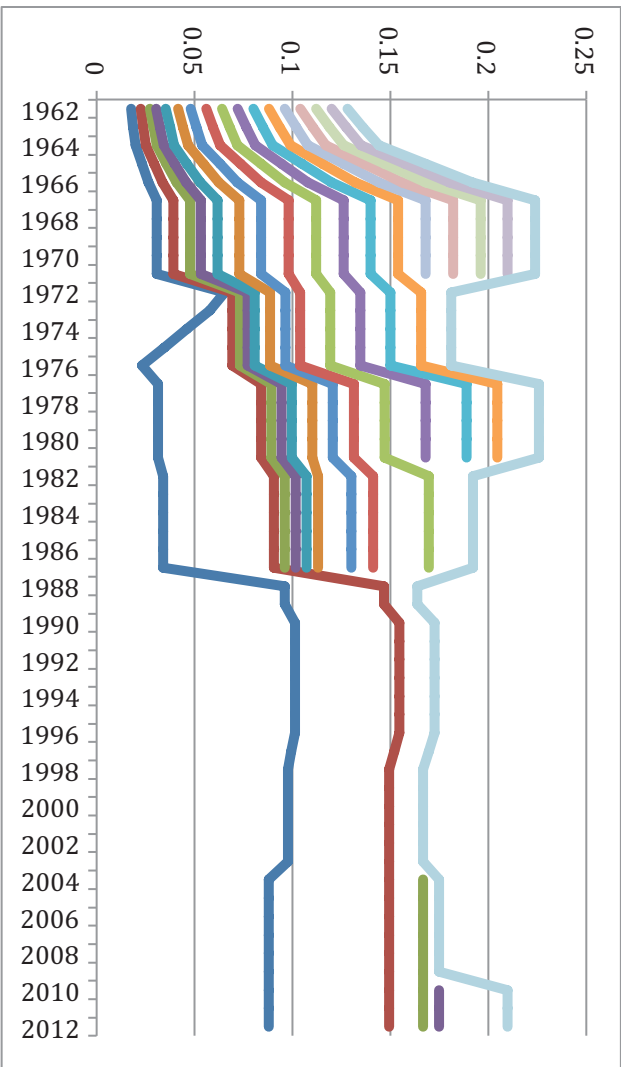
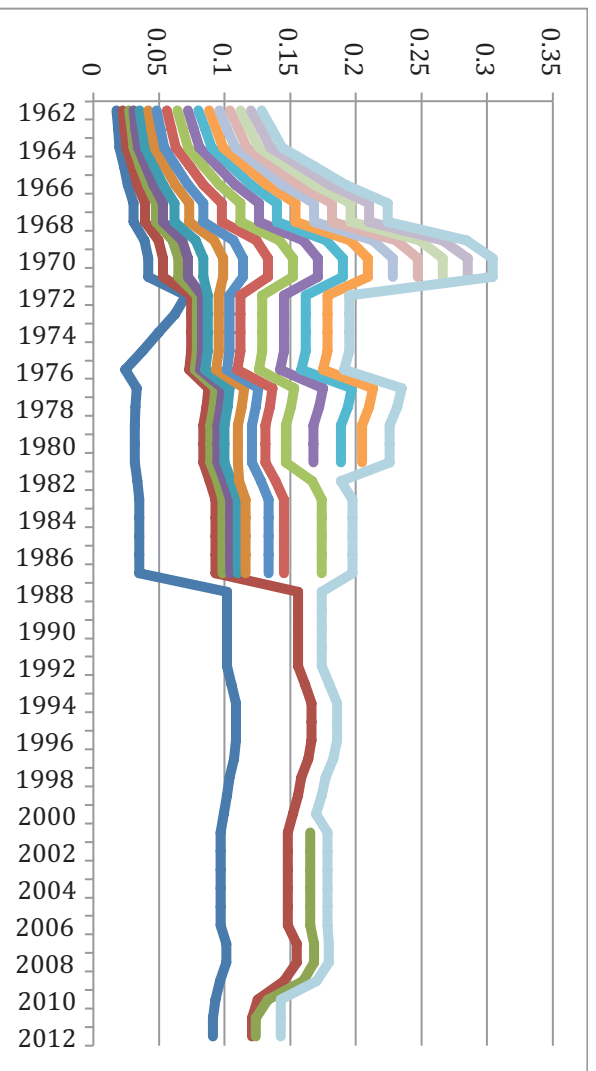


Table 3A
Nova Scotia Tax Rates, 1962-2012

Table 4A
New Brunswick Tax Rates, 1962-2012



ASB 2013
SFX, (NE)

Egbert McGraw, Université de Moncton,
(Canada).
Yves Robichaud, Université Laurentienne
(Canada).

ANALYSE DES ENJEUX FISCAUX

Les gouvernements fédéral et provinciaux font face à des dettes accumulées et des déficits importants et doivent tenter de rééquilibrer leurs budgets soit par une augmentation des revenus ou une diminution des dépenses. Dans ce travail, nous avons fait l'analyse de l'assiette fiscale de la province du Nouveau-Brunswick pour identifier les enjeux fiscaux et mieux comprendre les contraintes au redressement des finances publiques.

Le cas de la province du Nouveau-Brunswick

L'économie mondiale fait présentement face à des défis majeurs et plusieurs pays, dont la Grèce, ont dû ou devront adopter des mesures draconiennes pour équilibrer leurs budgets et rétablir les finances publiques. Le Canada a réussi à date à s'en sortir mais les gouvernements, tant fédéral que provinciaux font face à dettes accumulées et des déficits importants. Ils doivent donc tenter de rééquilibrer leurs budgets soit par une augmentation des revenus ou une diminution des dépenses.

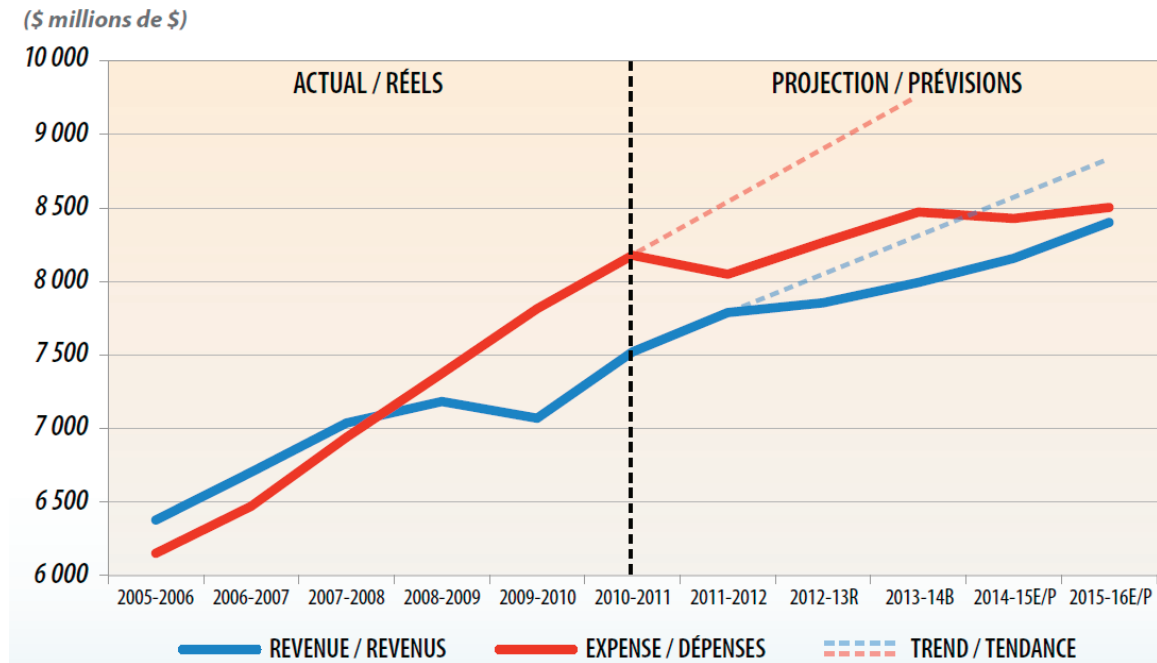
L'objectif de ce travail est d'analyser l'assiette fiscale de la province du Nouveau Brunswick afin d'identifier les enjeux fiscaux et de mieux comprendre les contraintes au redressement des finances publiques. Pour y parvenir, une analyse du budget de la province et des revenus des contribuables sera effectuée afin d'identifier en premier lieu, les sources de revenu provinciales et, en deuxième lieu l'apport de chacun des contribuables à l'assiette fiscale selon les différentes catégories de revenu. L'ensemble de ces analyses permettront de mieux comprendre les marges de manœuvre disponibles relativement à l'impôt sur le revenu des particuliers dans l'augmentation des revenus. A cette fin, les données les plus récentes publiées par la province du Nouveau Brunswick ainsi que les données de Statistique Canada disponibles à la date de l'étude seront utilisées

Le plan de cet article est le suivant : la première partie de l'article aborde la situation fiscale actuelle de la province du Nouveau-Brunswick en termes de dépenses et de revenus. A cet effet, les recettes de la province seront analysées. La deuxième partie portera sur l'analyse des revenus et des impôts personnels payés par les contribuables de cette province. Le tout sera suivi par une conclusion.

Situation fiscale du N-B en 2012-2013

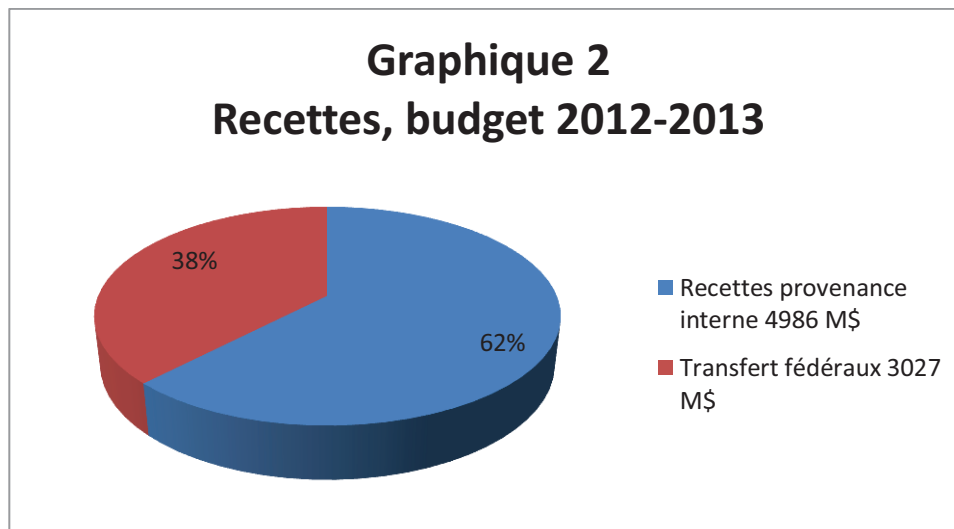
Tel que démontré au graphique 1, la province du Nouveau-Brunswick, tout comme les autres provinces canadiennes, fait face à une augmentation vertigineuse des dépenses. Même en essayant de contrôler les dépenses, la province fait face à un déficit de 182,9 M\$ pour l'année 2012-2013, et la tendance se maintiendra pour les prochaines années si des mesures ne sont pas prises pour corriger la situation. En plus de la rationalisation des dépenses, la Province doit aussi

faire une analyse de ses sources de revenus pour balancer le budget. La prochaine section va présenter les différentes sources de revenu de la Province.



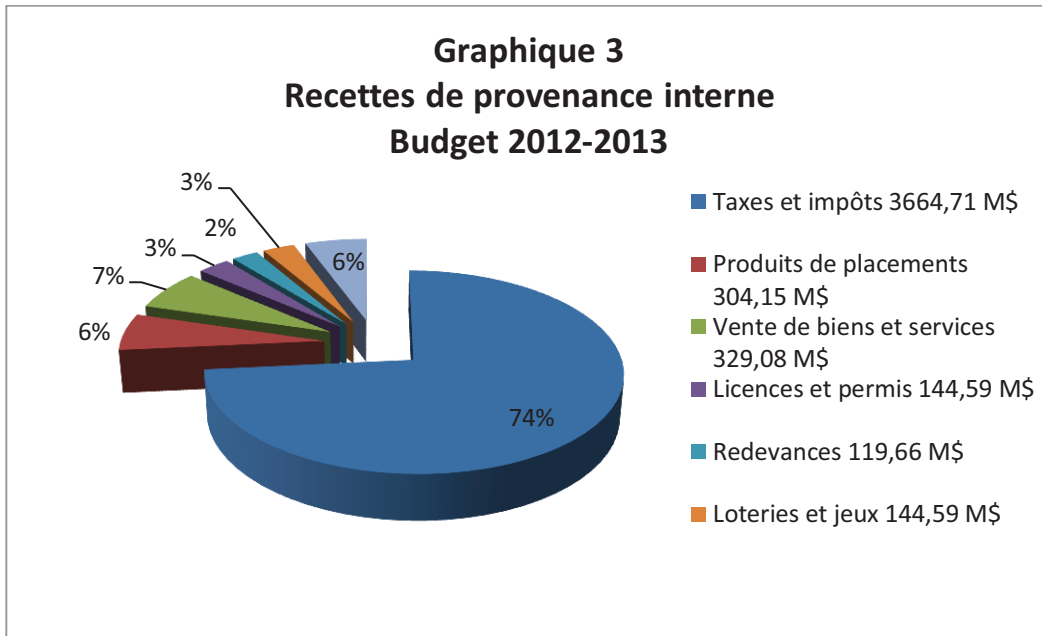
Source : Ministère des finances, Province du Nouveau Brunswick, Budget 2012-2013

Tel que présenté au graphique 2, les recettes provinciales proviennent de deux sources principales : les recettes de provenance interne qui sont évaluées à 4 986 millions de dollars (62,20%) et les paiements de transferts fédéraux qui représentent des recettes de 3 027 millions de dollars ou 37,8% des recettes provinciales. Ensemble, ces deux sources de revenu devraient atteindre 8 013 millions de dollars pour l'année financière 2012-2013.



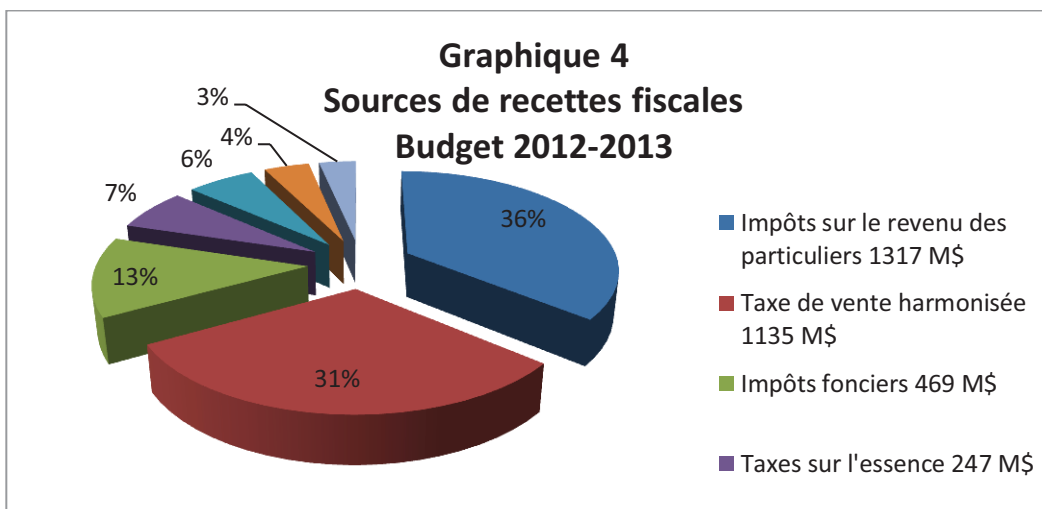
Source : ministère des Finances, Province du Nouveau Brunswick, Budget 2012-2013

Les recettes de provenance interne (graphique 3) comprennent les recettes fiscales (taxes et impôts, 74 %), les produits de placement (6 %), les ventes de biens et services (7 %), les licences et permis (3 %), les redevances découlant des ressources naturelles (2 %) et les recettes provenant des loteries et des jeux (3 %).



Source : Ministère des Finances, Province du Nouveau Brunswick, Budget 2012-2013

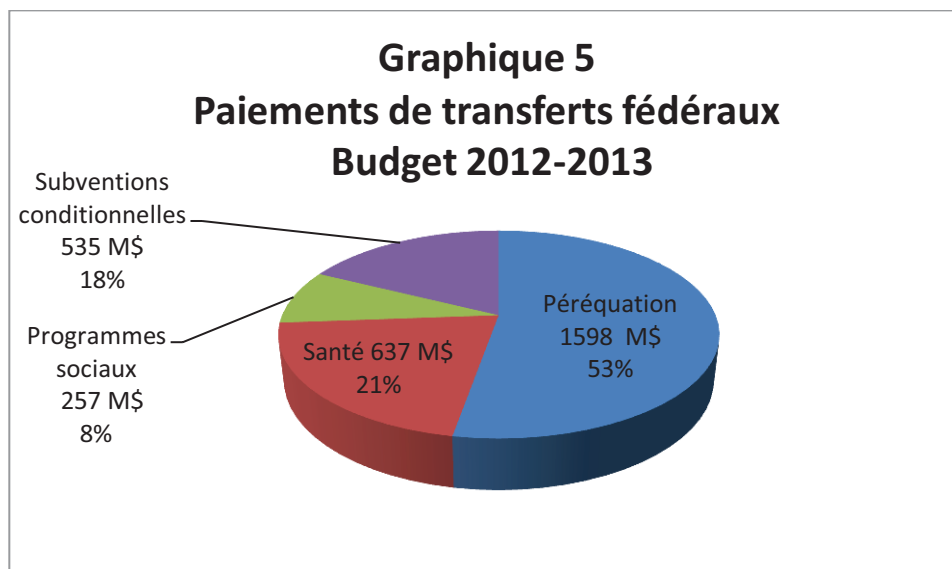
Le graphique 4 présente la source des taxes et impôts qui représentent la grande partie des recettes internes pour un total de 3 664 millions de dollars, soit 74 % des recettes totales. Les principales sources provenant de l'impôt sur le revenu des particuliers (36%), la taxe de vente harmonisée (TVH) (31%), l'impôt foncier (13%), la taxe sur l'essence et les carburants (7%), l'impôt sur le revenu des corporations (6%), la taxe sur le tabac (4%), et d'autres recettes (3%).



Source : Ministère des Finances, Province du Nouveau Brunswick, Budget 2012-2013.

Tel qu'indiqué précédemment, la deuxième source de recettes pour la province du Nouveau Brunswick provient des paiements de transferts fédéraux (3027 M\$) dont la péréquation, le transfert canadien en matière de santé, le transfert canadien en matière de programmes sociaux, et les subventions conditionnelles. Le graphique 5 présente la distribution de ces transferts fédéraux à la province du Nouveau Brunswick. Tel que démontré dans le graphique, le programme de péréquation fédéral est la principale source de recette de la province, qui est évalué à 1 598 millions de dollars, soit 19,9% des recettes totales de la province ou 53% des transferts fédéraux. La péréquation est un transfert inconditionnel du gouvernement fédéral et est de loin le plus important programme de transferts fédéraux à la province. Le but du programme de péréquation est de relever, à une norme établie, la capacité de générer des recettes par habitant des provinces bénéficiaires, dans le but d'offrir des services publics comparables à travers le Canada et de permettre un taux d'imposition comparable à tous les citoyens.

Les deux autres programmes de transfert fédéral sont le transfert canadien en matière de santé (637 M\$) et le programme de transfert canadien en matière de programmes sociaux (257 M\$) qui représentent respectivement 21% et 8% des transferts fédéraux en environ 11% des recettes de la province pour l'année 2012-2013. Ces recettes appuient les soins de santé, l'enseignement post secondaire, l'aide sociale et les services sociaux et autres programmes délivrés par la province. Les subventions conditionnelles du Canada représentent 535 millions de dollars, soit 18% des transferts fédéraux.



Source : Ministère des Finances, Province du Nouveau Brunswick, Budget 2012-2013.

L'analyse des recettes de la province du Nouveau Brunswick a démontré que les transferts fédéraux représentent 38% des recettes provinciales, et que les impôts sur le revenu des particuliers (1 317 M\$) et la taxe de vente harmonisée (1 135 M\$) représentent les deux principales sources de revenu interne, soit 49% des recettes de provenance interne ou 67% des taxes et impôts. Il est aussi important de noter que la taxe sur l'essence, les impôts sur le revenu des corporations ainsi que les taxes sur le tabac ne représentent que 7%, 6% et 4% des taxes et impôts respectivement.

Analyse - impôt sur le revenu des particuliers

Dans la section précédente nous avons démontré que l'impôt sur le revenu des particuliers représente la première source de recettes de provenance interne de la province. Dans la partie qui suit, nous allons faire une analyse des revenus et des impôts payés par les contribuables de la province du Nouveau Brunswick. Au besoin, nous comparerons ces données avec celles du Canada et d'autres provinces. A cet effet, le tableau 1 présente les impôts directs en provenance des particuliers pour la province du Nouveau Brunswick tant au niveau fédéral que provincial pour la période 2001 à 2009. Les données démontrent que l'impôt direct des particuliers tant au niveau provincial que fédéral de la période de 2001 à 2009 a augmenté de 11,65% au fédéral (1785 M\$ vs 1577 M\$) et de 26% au provincial (1248 M\$ vs 923 M\$). Pour la province du Nouveau Brunswick, nous pouvons observer une diminution des impôts des individus pour l'année 2009 par rapport à 2008, ce qui pourrait s'expliquer par la diminution des taux d'impôt au Nouveau Brunswick au cours de cette période. Il faut aussi signaler que les impôts prévus par la province du Nouveau Brunswick pour l'année 2012 à 2013 (1 317 M\$) sont inférieurs au niveau de 2008 (1 339 M\$), ce qui semble donner raison aux personnes qui, à l'époque, estimaient que la province n'avait pas la capacité financière de réduire les taux d'impôts des individus. Il est à remarquer que la province a rétabli, lors du dernier budget en 2013, les taux d'impôts comme ils l'étaient il y a quelques années.

Tableau 1
Impôts directs des particuliers pour la province du Nouveau-Brunswick
(CANSIM 384-0006) de 2001 à 2009

	2001	2002	2003	2004	2005	2006	2007	2008	2009
Total impôts directs (particuliers) (M\$)	2,500	2,480	2,494	2,613	2,713	2,766	3,000	3,202	3,033
Fédéral (M\$)	1,577	1,514	1,527	1,599	1,634	1,636	1,742	1,863	1,785
Provincial (M\$)	923	966	967	1,014	1,079	1,130	1,258	1,339	1,248

Le tableau 2 fait ressortir le nombre de contribuables de la province du Nouveau Brunswick selon les différentes catégories de revenu. Les contribuables représentent les personnes qui ont produit une déclaration d'impôt pour la période de référence et qui étaient vivants à la fin de l'année. On observe une légère augmentation du nombre de contribuables au cours des années; il y avait 591 500 déclarants en 2010, comparativement à 574 070 déclarants pour 2005, soit une augmentation de 3%. Par contre, le nombre de contribuables a diminué pendant les 10 dernières années. Les statistiques montrent une variation négative ou nulle du nombre de résidents pendant la période de 2001 à 2010.

En comparaison, les statistiques canadiennes (tableau 3) montrent une progression (variation en pourcentage depuis 5 ans) plus grande du nombre de contribuables pendant cette période soit une progression qui varie entre 6% et 10% comparativement à une progression de 3% à 6% pour la province du Nouveau Brunswick (tableau 2). Il en est de même pour la progression du nombre de canadiens qui varie de 4% à 5% pendant cette période comparativement à la province du Nouveau Brunswick qui montre une variation négative pendant plusieurs périodes de 5 ans.

Le tableau 2 présente aussi le pourcentage de déclarants avec un revenu supérieur à un certain seuil. À titre d'exemple, pour l'année 2010, 3% des déclarants du Nouveau Brunswick avaient un revenu total supérieur à 100 000\$, 8% un revenu total supérieur à 75 000\$, 20% un revenu supérieur à 50 000\$, 37% supérieurs à 35 000\$, 53% supérieurs à 25 000 et 74% supérieurs à 15 000\$. Quoique le pourcentage d'individus dans les catégories de revenus supérieures a légèrement augmenté pendant ces 10 dernières années, le pourcentage de déclarants dans des catégories supérieures est largement inférieur aux statistiques canadiennes (tableau 3) avec 6% des répondants avec un revenu total de 100 000\$ et plus, et 12% avec un revenu de 75 000\$ et plus. De plus, bien que le revenu médian des déclarants du N.B. est passé de 19 000\$ en 2001 à 26 610\$ en 2010 (tableau 2), soit une augmentation de 28,6%, il est quand même 9% inférieur à celui des canadiens qui est de 29 250 \$ en 2010 (tableau 3).

Une autre statistique alarmante pour la province du Nouveau Brunswick est le pourcentage de déclarants qui ont reçu de l'assurance emploi qui est toujours supérieur (presque le double) aux statistiques canadiennes et cela pour chacune des années soit de 2001 à 2009. À titre d'exemple, en 2010, au N.B. il était de 26% comparativement à 15% pour le Canada.

Nous pouvons aussi nous poser des questions sur la contribution fiscale des mieux nantis. À cet effet, les tableaux 4 et 5 présentent la part des revenus ainsi que la part de l'impôt sur le revenu par quintile. Les contribuables sont classifiés en 5 quintiles représentant chacun la part des revenus de 20% des contribuables. Ainsi le dernier quintile représente la part des contribuables avec le revenu le plus élevé soit celle entre 80% et 100%. Il en est de même pour la classification de la part de l'impôt sur le revenu payée par les contribuables où chacun des 5 quintiles représente 20% des contributions en impôt des déclarants. Cette analyse se fera pour 2 types de familles économiques, soit une famille de deux personnes ou plus (tableau 4), ainsi que pour les personnes seules (tableau 5).

Le tableau 4-A présente la part du revenu en pourcentage d'une famille économique de deux personnes ou plus pour la province du Nouveau Brunswick. À titre d'exemple, pour l'année 2011, les déclarants dans le quintile supérieur gagnent 46% des revenus, tandis que la part du revenu des déclarants dans le quintile inférieur gagne 2,2% des revenus. Le quatrième quintile gagne 25,8% des revenus, le troisième quintile, 16,5%, et le deuxième quintile, 9,5%. En combinant ces quintiles, les 40% des plus hauts salariés gagnent 71,8% (25,8% + 46%) des revenus. Ces pourcentages de la part des revenus par les différentes catégories de revenu ont peu changé pendant les années 2001 à 2011.

Le tableau 4-B présente leur part de l'impôt sur le revenu par quintile. Pour 2011, les déclarants dans le quintile supérieur contribuent 55% de l'impôt sur le revenu, tandis que la contribution à l'impôt sur le revenu des déclarants dans le quintile inférieur est de 0,08%. À titre de comparaison, en combinant ces quintiles, les 40% des déclarants les plus imposés contribuent 79,8% (24,8% + 55%) de l'impôt sur le revenu, ou pour résumer, 40% des familles contribuent 80% à l'impôt sur le revenu et le reste des cotisants, soit 60% des familles contribuent environ 20% aux recettes fiscales relié à l'impôt sur le revenu.

Le tableau 5-A présente la part du revenu en pourcentage d'une personne seule pour la province du Nouveau Brunswick. À titre d'exemple, pour l'année 2011, les déclarants dans le quintile

supérieur gagnent 52,3% des revenus, tandis que la part du revenu des déclarants dans le quintile inférieur est 0% des revenus. Le quatrième quintile gagne 27,3% des revenus, etc. En combinant ces quintiles, les 40% des cotisants avec les revenus plus élevés gagnent 79,6% (27,3% + 52,3%) des revenus et les 60% des déclarants avec les plus bas revenus gagnent 20,4% des revenus.

Le tableau 5-B présente leur part de l'impôt sur le revenu par quintile. Pour 2011, les déclarants dans le quintile supérieur contribuent 60,5% de l'impôt sur le revenu, tandis que la contribution à l'impôt sur le revenu des déclarants dans le quintile inférieur est de 1,9%. À titre de comparaison, en combinant ces quintiles, les 40% des déclarants les plus imposés contribuent 87,2% (26,7% + 60,5%) à l'impôt sur le revenu, tandis que 60% des déclarants les plus pauvres contribuent 12,8% à l'impôt sur le revenu. Les statistiques démontrent que de 2001 à 2011, la part du revenu du quintile supérieur ainsi que sa part de l'impôt diminue au cours des années.

De façon générale, ces tableaux démontrent que les mieux nantis paient une proportion d'impôt supérieur à leur part de revenu, ce qui peut s'expliquer par le fait que nous avons un système d'impôt progressif, et que le taux d'impôt augmente avec l'augmentation des revenus. De plus, les personnes seules paient un pourcentage plus élevé que des déclarants d'une famille économique de deux personnes ou plus.

Par contre, il y a lieu de comparer la contribution à l'impôt des déclarants du Nouveau Brunswick par rapport à l'Atlantique ainsi que le Canada. Pour simplifier l'analyse nous comparerons uniquement les déclarants dans le quintile supérieur parce que tel que démontré, les déclarants dans les quintiles inférieurs paient peu d'impôt et socialement il serait plus acceptable d'analyser davantage la contribution des mieux nantis. Notre analyse se fera pour deux types de familles économiques, les familles de deux personnes et plus, et pour les personnes seules.

Le tableau 4c présente l'impôt moyen sur le revenu pour les familles économiques de deux personnes ou plus dans le quintile supérieur. Pour l'année 2011, les familles dans le quintile supérieur payaient un impôt moyen de 30,700\$ comparativement à 34,300\$ pour l'Atlantique et de 44,200\$ pour le Canada. Les déclarants du Nouveau-Brunswick ont aussi un impôt moyen pendant les années 2001 à 2011. Le tableau 4d présente le taux implicite d'impôt pour ces mêmes familles. Pour 2011, les familles dans le quintile supérieur avaient un taux implicite d'impôt de 19,5% comparativement à 20,7% pour l'Atlantique et 21,9% pour le Canada.

L'analyse du taux moyen d'impôt sur le revenu et du taux implicite d'impôt pour une personne seule présente des résultats similaires pour les déclarants du Nouveau-Brunswick. Les résultats sont présentés aux tableaux 5c et 5d. Un déclarant seul du Nouveau-Brunswick paye un impôt moyen de 11,400\$ comparativement à 12,900\$ pour l'Atlantique et de 18,600\$ pour le Canada. Le taux implicite d'impôt est de 17,8%, de 19,0% et de 21,6% respectivement, démontant que pour 2011, ainsi que pour les années de l'étude, un déclarant seul du Nouveau-Brunswick, avait un taux implicite d'impôt inférieur que l'Atlantique ainsi que le Canada.

L'impôt moyen sur le revenu ainsi que le taux implicite d'impôt inférieur pour les déclarants du Nouveau-Brunswick dans le quintile supérieur peut s'expliquer en partie parce qu'ils ont des revenus inférieurs que ceux de l'Atlantique ainsi que du Canada. Le taux d'imposition pourrait être une autre raison. Le tableau 6 présente le taux d'imposition provincial et le taux combiné

(provincial et fédéral) pour les provinces et territoires canadiens pour 2011. La première colonne présente le taux marginal provincial, la deuxième colonne le taux combiné (provincial et fédéral), la troisième colonne représente le taux marginal pour un gain en capital et les deux dernières le taux d'imposition pour les dividendes déterminés et non déterminés. Le tableau 6 démontre que le taux provincial pour le Nouveau Brunswick était de 14,3% (43,30% combiné), soit le septième plus bas taux provincial après le Québec (24 %, 48,22% combiné), la Nouvelle Écosse (21%, 50% combiné), le Manitoba (17,4%, 46,4% combiné), l'Île du Prince Édouard (16,7%, 47,37% combiné), la Saskatchewan (15%, 44% combiné), et la Colombie Britannique (14,7%, 43,7% combiné). Ce taux d'imposition moindre de la province du Nouveau Brunswick pourrait ajouter à l'explication du taux implicite d'impôt du Nouveau Brunswick. ¹

Les tableaux 7 et 8 présentent le revenu moyen par quintile pour le Nouveau Brunswick, l'Atlantique, ainsi que le Canada pour une famille économique de deux personnes ou plus et pour une personne seule pour les années 2001 à 2011. Le tableau 7-A démontre que, pour l'année 2011, le revenu moyen pour les familles dans le quintile supérieur était de 152 500\$ comparativement à 7 300\$ pour le quintile inférieur ce qui démontre une large disparité dans le revenu des familles, même s'il y a eu progression du revenu moyen pendant ces années, à raison de 31% pour le quintile inférieur (7300\$ vs 5000\$) et de 11% pour le quintile supérieur (152500\$ vs 135 700\$) pour une famille économique de deux personnes. Par contre, même si le pourcentage d'augmentation du revenu moyen est supérieur pour le quintile inférieur par rapport au quintile supérieur, la différence absolue est troublante avec une augmentation du revenu de 2300\$ pendant 10 ans pour les moins fortunés comparativement à 16800\$ pour le quintile supérieur.

Le tableau 7-B et 7-C présente le revenu moyen par quintile pour l'Atlantique ainsi que le Canada. Les statistiques démontrent que le revenu moyen du quartile supérieur pour 2011 en Atlantique était supérieur de 5% (152 500\$ vs 160 600\$) à celui du Nouveau-Brunswick, et de 23% (152500\$ vs 198 000\$) pour le Canada. L'ensemble des déclarants du Nouveau Brunswick ont un revenu moindre qu'ailleurs au pays, et ce pendant la période étudiée.

Il en est de même pour les statistiques pour une personne seule tel que démontré au tableau 8. Le tableau 8-A démontre que, pour l'année 2011, le revenu moyen pour les personnes seules dans le quintile supérieur était de 61 000\$ comparativement à un revenu moyen nul pour le quintile inférieur ce qui démontre encore une grande disparité au niveau des individus. Le tableau 8-B et 8-C démontrent que le revenu moyen du quintile supérieur pour 2011 en Atlantique était supérieur de 6% à celui du Nouveau-Brunswick, et de 27% pour le Canada. L'ensemble des déclarants seuls du Nouveau Brunswick ont un revenu moyen moindre qu'ailleurs au pays, et ce pendant la période étudiée. Les statistiques des tableaux 7 et 8 démontrent donc que les familles ainsi que les individus du Nouveau Brunswick ont un revenu moyen inférieur à celui des familles

¹ Nous tenons à mettre en garde le lecteur que le système fiscal canadien en plus d'utiliser des taux d'imposition progressifs utilise des crédits non remboursable qui peuvent différer d'une province à l'autre, et une comparaison uniquement du taux marginal d'impôt ne donne pas un portrait fidèle du fardeau fiscal pour un individu. Il faudrait faire une analyse cas par cas pour évaluer adéquatement la situation fiscale d'un contribuable, d'où l'utilisation d'une moyenne par quintile.

et des individus du Canada et de l'Atlantique ce qui limite la capacité de la province d'augmenter les recettes provenant de l'impôt sur le revenu des particuliers.

Conclusion

La première réaction dans les discours populaires est d'augmenter les impôts des plus riches, parce que chaque contribuable juge qu'il paye trop d'impôts. Il suffit d'aller prendre un café dans un café populaire pour entendre tout le monde se lamenter à l'effet qu'ils sont trop imposés, et même que le gouvernement les égorge, pour utiliser leurs expressions. L'analyse des données a démontré que la province du Nouveau-Brunswick est dans une situation précaire pour augmenter les recettes fiscales par l'entremise de l'impôt sur le revenu, environ 40% des contribuables de la province payent déjà au-delà de 80% des impôts. Les capacités d'imposition sont aussi limitées parce que 47% des déclarants ont un revenu inférieur à 25 000 \$, et le revenu moyen est inférieur à l'ensemble du Canada ainsi que de l'Atlantique. Du côté des recettes, le Nouveau-Brunswick est très dépendant des transferts fédéraux représentant 38% des recettes provinciales et la position du Fédéral de diminuer les transferts relié à la santé pourra avoir un impact sur le budget de la province. Peut-être la stratégie de diminuer les impôts personnels dans le but d'attirer plus de personnes ou d'entreprises n'a pas porté ses fruits, mais l'augmentation des taux ne saura pas redresser à elle seule la situation fiscale de la province.

Références :

Budget 2012-2013, Rebâtir le Nouveau-Brunswick ensemble, Ministère des finances, Province du Nouveau-Brunswick, 2012; www.gnb.ca/0024/index-f.asb

Fiche d'information-Finances, Recettes, Budget de 2012-2013, Ministère des Finances, Province du Nouveau Brunswick; <http://www.gnb.ca/0024/index-f.asb>

Planification fiscale personnelle, 2011-2012, Le reper, Association des comptables généraux accrédités du Canada, 2011.

Statistique Canada, CANSIM, Tableau 111-0004.

Statistique Canada, CANSIM, Tableau 202-0501.

Statistique Canada, CANSIM, Tableau 202-0701.

Statistique Canada, CANSIM, Tableau 384-0006.

www.gnb.ca/0024/index-f.asb

Tableau 2
 Nombre de déclarants par catégorie de revenu pour la province du Nouveau-Brunswick
 de 2001 à 2010 (CANSIM 111-0004)

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nombre de déclarants	565 410	567 700	569 690	572 490	574 070	575 060	580 740	585 060	589 060	591 500
Variation en % depuis 5 ans	6	6	5	4	3	2	2	3	3	3
Résidents	739 160	739 930	737 990	737 800	730 570	729 840	730 630	732 390	735 680	737 220
Variation en % depuis 5 ans	-1	0	0	0	-1	-1	-1	-1	0	0
Pourcentage avec un revenu total :										
-15 000 \$ et plus	60	62	63	65	66	69	71	72	73	74
-25 000 \$ et plus	38	40	41	43	44	46	49	51	52	53
-35 000 \$ et plus	23	24	25	26	28	30	32	35	35	37
-50 000 \$ et plus	10	11	12	13	14	16	17	19	20	20
-75 000 \$ et plus	3	3	4	4	4	5	6	7	7	8
-100 000 \$ et plus	1	1	1	2	2	2	2	3	3	3
Revenu médian	19 000	19 700	20 300	21 000	21 800	22 900	24 270	25 350	25 890	26 610
Revenu médian (indice canadien)	84,1	85,3	86,0	86,1	85,8	86,4	86,8	87,7	89,8	91,0
Pourcentage avec assurance-emploi	27	27	27	27	26	25	25	24	26	26

Tableau 3
 Nombre de déclarants par catégorie de revenu pour le Canada
 de 2001 à 2010 (CANSIM 111-0004)

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nombre de déclarants	22 804 290	22 967 720	23 267 830	23 624 530	23 951 820	24 258 900	24 623 550	24 986 960	25 244 320	25 484 240
Variation en % depuis 5 ans	10	9	9	8	8	6				
Canadiens	30 228 420	30 465 870	30 695 230	31 041 180	31 099 150	31 492 030	31 772 910	32 124 240	32 425 050	32 700 430
Variation en % depuis 5 ans	5	5	5	5	4	4				
Pourcentage avec un revenu total :										
-15 000 \$ et plus	65	66	67	68	69	71	73	74	74	74
-25 000 \$ et plus	46	47	48	49	51	52	54	56	55	56
-35 000 \$ et plus	31	32	33	35	36	38	40	42	42	43
-50 000 \$ et plus	17	18	19	20	21	23	25	26	26	27
-75 000 \$ et plus	6	6	7	8	9	10	11	12	12	12
-100 000 \$ et plus	3	3	3	3	4	4	5	5	5	6
Revenu médian	22 600	23 100	23 600	24 400	25 400	26 500	27 960	28 920	28 840	29 250
Revenu médian (indice canadien)	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Pourcentage avec assurance-emploi	14	14	14	14	13	13	13	13	16	15

Tableau 4

Part du revenu (pourcentage) d'une famille économique deux personnes ou plus pour la province du Nouveau-Brunswick (202-0701)

A

de 2001 à 2011 (CANSIM 202-0701)

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Quintile inférieur	1,7	1,7	1,7	1,9	1,8	2,0	2,1	2,3	2,2	2,4	2,2
Deuxième quintile	9,5	9,2	9,3	9,5	9,2	9,4	9,1	9,7	9,3	9,4	9,5
Troisième quintile	17,0	17,0	16,7	16,5	16,6	16,2	16,4	17,3	16,6	17,0	16,5
Quatrième quintile	25,5	25,8	25,2	25,4	25,7	25,4	25,5	25,8	25,8	25,6	25,8
Quintile supérieur	46,4	46,2	47,1	46,7	46,8	47,0	46,9	44,9	46,0	45,6	46,0

Part de l'impôt sur le revenu d'une famille économique deux personnes ou plus pour la province du Nouveau-Brunswick (202-0501)

B

de 2001 à 2011 (CANSIM 202-0501)

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Quintile inférieur	1,1	0,9	1,5	1,3	1,9	1,4	1,1	0,8	0,8	4,2	0,8
Deuxième quintile	6,7	6,3	6,5	6,8	6,3	6,2	5,5	6,5	4,8	4,9	6,3
Troisième quintile	14,5	14,7	14,1	14,4	13,4	12,9	12,7	13,8	13,3	14,1	13,1
Quatrième quintile	24,0	24,2	23,3	23,0	24,0	24,4	23,8	25,0	25,6	23,9	24,8
Quintile supérieur	53,8	53,9	54,5	54,5	54,4	55,0	56,8	54,0	55,5	52,8	55,0

Impôt moyen sur le revenu d'une famille de deux personnes ou plus dans le quintile supérieur pour le Nouveau-Brunswick, les provinces de l'Atlantique et le Canada pour la période de 2001 à 2011 (CANSIM 202-0501)

C

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Nouveau-Brunswick	30 200	30 100	31 300	31 400	28 900	30 300	31 900	30 200	30 700	30 000	30 700
Atlantique	31 100	31 900	30 300	30 800	31 600	33 200	34 400	34 700	33 700	33 900	34 300
Canada	42 900	42 500	41 700	43 100	42 300	42 600	43 700	46 200	43 200	43 200	44 200

Taux implicite d'impôt sur le revenu (pourcentage) d'une famille de deux personnes ou plus dans le quintile supérieur pour le Nouveau-Brunswick, les provinces de l'Atlantique et le Canada pour la période de 2001 à 2011 (CANSIM 202-0501)

D

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Nouveau-Brunswick	21,6	21,5	22,0	21,7	20,5	20,8	21,0	20,7	19,8	19,5	19,5
Atlantique	21,9	22,1	21,7	21,7	21,5	21,8	21,9	21,9	20,7	20,8	20,7
Canada	23,4	23,2	23,2	23,2	22,8	22,5	22,1	22,7	21,6	21,5	21,9

Tableau 5

Part du revenu (pourcentage) pour une personne seule pour la province du Nouveau-Brunswick
de 2001 à 2011 (CANSIM 202-0701)

A

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Quintile inférieur	0,0	0,0	-0,3	-0,1	-0,1	0,0	0,0	-0,7	-0,4	-0,1	0,0
Deuxième quintile	2,8	2,3	1,9	2,6	3,1	3,3	4,2	3,0	4,4	3,9	5,2
Troisième quintile	10,4	11,4	11,9	13,3	12,4	13,3	13,9	12,9	14,5	14,4	15,2
Quatrième quintile	22,9	26,1	27,5	27,4	27,9	27,6	25,6	26,9	26,7	26,1	27,3
Quintile supérieur	63,9	60,2	58,9	56,8	56,6	55,7	56,4	57,9	54,8	55,8	52,3

Part de l'impôt sur le revenu d'une personne seule pour la province du Nouveau-Brunswick
de 2001 à 2011 (CANSIM 202-0501)

B

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Quintile inférieur	2,9	0,4	0,5	0,5	0,6	3,4	0,4	0,0	0,0	1,5	1,9
Deuxième quintile	1,2	2,6	1,9	3,0	1,5	1,5	2,8	1,0	2,0	1,7	3,0
Troisième quintile	5,4	6,3	7,9	9,5	6,8	7,7	7,3	7,2	10,0	7,4	8,0
Quatrième quintile	17,3	20,6	22,3	22,7	24,4	23,7	20,7	22,5	24,0	22,9	26,7
Quintile supérieur	73,2	70,0	67,3	64,3	66,7	63,7	68,7	69,2	64,1	66,5	60,5

Impôt moyen sur le revenu pour une personne seule dans le quintile supérieur pour le Nouveau-Brunswick,
les provinces de l'Atlantique et le Canada pour la période de 2001 à 2011 (CANSIM 202-0501)

C

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Nouveau-Brunswick	17 000	11 400	11 100	10 300	11 100	11 800	14 100	15 100	12 100	13 400	11 400
Atlantique	14 000	12 000	14 000	13 000	12 800	13 100	14 400	15 200	13 200	14 200	12 900
Canada	18 500	18 600	20 100	19 600	19 900	21 200	21 100	19 500	19 600	21 500	18 600

Taux implicite d'impôt sur le revenu (pourcentage) pour une personne seule dans le quintile supérieur pour le
Nouveau-Brunswick, les provinces de l'Atlantique et le Canada pour la période de 2001 à 2011 (CANSIM 202-0501)

D

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Nouveau-Brunswick	23,6	20,0	19,6	19,1	18,9	19,8	20,8	21,7	18,1	19,8	17,8
Atlantique	22,0	20,6	22,0	20,8	20,6	20,6	21,0	21,8	19,1	20,4	19,0
Canada	22,9	22,8	23,6	23,3	23,0	23,3	23,1	21,8	21,6	23,0	21,6

Tableau 6
Taux d'imposition combinés (fédéral et provincial/territorial) maximaux - 2011

Province/Territoire	Taux provincial/ territorial	Revenu ordinaire	Gains en capital	Dividendes déterminés	Dividendes non déterminés
Alberta	10,00 %	39,00 %	19,50 %	17,72 %	27,71 %
Colombie-Britannique	14,70	43,70	21,85	23,91	33,71
Île-du-Prince-Édouard	16,70	47,37	23,69	27,33	41,17
Manitoba	17,40	46,40	23,20	26,74	39,19
Nouveau-Brunswick	14,30	43,30	21,65	20,96	30,83
Nouvelle-Écosse	21,00	50,00	25,00	34,85	36,21
Nunavut	11,50	40,50	20,25	25,72	28,96
Ontario	11,16	46,41	23,20	28,19	32,57
Québec	24,00	48,22	24,11	31,85	36,35
Saskatchewan	15,00	44,00	22,00	23,36	32,08
Terre-Neuve-et-Labrador	13,30	42,30	21,15	20,96	29,96
Territoires-du-Nord-Ouest	14,05	43,05	21,53	21,31	29,65
Yukon	12,76	42,40	21,20	14,40	30,41

Tableau 7
Revenu moyen pour une famille économique de deux personnes ou plus (CANSIM 202-0701)
de 2001 à 2011

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
A Nouveau-Brunswick											
Quintile inférieur*	5 000	4 900	4 900	5 800	5 200	6 100	6 700	7 300	7 100	7 900	7 300
Deuxième quintile	27 800	27 000	27 500	28 400	27 000	28 400	28 500	30 700	30 300	30 900	31 400
Troisième quintile	49 600	49 800	49 000	49 600	48 600	48 900	51 500	54 900	54 400	56 300	54 600
Quatrième quintile	74 500	75 300	74 300	76 300	75 400	76 800	80 200	82 000	84 400	84 400	85 400
Quintile supérieur	135 700	135 100	138 600	140 500	137 400	141 800	147 300	142 700	150 400	150 600	152 500
B Provinces de l'Atlantique											
Quintile inférieur	4 500	4 600	4 500	5 300	6 100	6 900	7 300	7 100	6 800	6 800	7 700
Deuxième quintile	25 400	25 800	25 900	27 700	27 900	29 400	30 100	30 500	29 900	30 300	32 500
Troisième quintile	48 000	48 200	47 700	49 400	49 500	51 100	53 300	53 800	53 500	54 800	57 000
Quatrième quintile	74 500	75 000	74 600	76 300	76 200	78 400	82 000	82 800	85 100	85 700	89 100
Quintile supérieur	137 100	139 800	135 400	137 300	142 600	147 600	152 100	154 300	157 100	158 600	160 600
C Canada											
Quintile inférieur	9 500	9 800	9 900	10 300	10 600	11 600	12 400	11 700	10 500	10 400	11 500
Deuxième quintile	37 400	37 300	37 000	37 600	38 700	40 000	41 300	41 100	38 700	38 600	40 500
Troisième quintile	63 500	62 800	62 800	63 800	64 700	66 000	67 800	68 700	66 900	67 000	68 500
Quatrième quintile	93 200	93 200	93 000	94 600	96 600	97 800	100 900	102 000	100 800	101 100	103 600
Quintile supérieur	180 200	180 000	176 900	183 100	183 000	186 200	194 400	199 700	196 200	196 400	198 000

Tableau 8
Revenu moyen (dollars) pour une personne seule (CANSIM 202-0701)
de 2001 à 2011

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
A Nouveau-Brunswick											
Quintile inférieur	0	0	-300	-100	-100	0	0	-800	-500	-200	0
Deuxième quintile	3 000	2 000	1 800	2 300	3 100	3 500	4 800	3 500	5 100	4 500	6 100
Troisième quintile	11 300	10 100	10 900	11 900	12 400	13 900	16 100	14 700	16 900	16 800	17 700
Quatrième quintile	24 700	23 300	25 100	24 600	27 700	28 800	29 600	30 800	31 100	30 400	31 900
Quintile supérieur	68 900	53 700	53 800	51 100	56 300	58 100	65 400	66 200	63 900	64 900	61 000
B Provinces de l'Atlantique											
Quintile inférieur	-300	0	-100	-200	-300	-100	-100	-300	-200	-100	0
Deuxième quintile	2 500	2 300	1 900	2 100	2 900	3 600	4 300	3 500	4 000	3 400	4 000
Troisième quintile	11 200	12 000	11 900	11 800	13 000	13 800	15 300	14 300	15 800	15 300	16 100
Quatrième quintile	25 100	26 300	27 100	26 400	28 800	29 900	30 000	30 800	31 300	31 800	31 700
Quintile supérieur	60 600	54 900	61 100	59 300	59 500	61 500	65 300	67 500	66 700	67 300	65 200
C Canada											
Quintile inférieur	-200	-100	0	0	0	100	200	200	0	0	0
Deuxième quintile	6 000	6 700	6 800	7 100	7 200	7 800	8 300	8 400	8 100	7 400	7 200
Troisième quintile	19 100	20 000	19 800	20 100	20 300	20 900	22 100	22 700	22 100	21 900	21 000
Quatrième quintile	36 900	37 300	37 800	38 000	38 500	39 600	40 300	41 900	41 200	40 900	39 100
Quintile supérieur	78 700	79 500	83 400	82 100	85 100	89 300	89 700	87 700	88 500	91 400	83 600

ASB 2013
Antigonish, Nova Scotia

Joseph Abekah
Sajedah Beykaei
Abdur Rahim
University of New Brunswick, Fredericton

OPERATIONALIZING COST-VOLUME-PROFIT ANALYSIS UNDER UNCERTAINTY

Cost-Volume-Profit Analysis is a major business planning tool. It is limited by its assumptions about known price and costs. This study uses MAPLE software to operationalize CVP under uncertainty for practical application. Illustrated examples show that the impact of expected selling price and expected variable cost on expected breakeven quantity is more sensitive than the impact of the standard deviations of selling price and variable cost. Specifically, when expected selling price decreases, the expected breakeven quantity increases rapidly. However, as the standard deviations of selling price or variable cost increases, expected breakeven quantity increases at the same rate.

1. Introduction

Minimizing costs to maximize profits is a management concern in all business productive efforts. Prior to proceeding with the production of a good or service management determines whether the revenues from the good or service to be produced will cover the costs that will be necessary for the productive efforts. These costs may be fixed or variable. If managers do not have accurate information about costs they will encounter difficulties in making production, pricing, and other related decisions. Since what is produced has to be sold before revenue can be earned, the company's profitability depends on (1) costs, and (2) sales quantity. Managers use Cost-Volume-Profit (CVP) analysis as a tool to understand the relationships between costs, volume, price, and profits.

The main objective of this paper is to demonstrate how C-V-P concepts can be adapted to the overall goal of maximizing expected profits. To this end, various sensitivity analyses based on the stochastic C-V-P model proposed by Hilliard and Leitch (1975, 1979) are conducted.

The rest of this paper is organized as follows: Section 2 explains the objectives of C-V-P analysis. Section 3 reviews C-V-P analysis under uncertainty. Section 4 describes the stochastic C-V-P analysis model development and performs sensitivity analyses of stochastic breakeven point. It is followed by illustrative numerical results from the model's stochastic breakeven point. Finally, contributions and limitations are provided in section 5.

2. Objectives of Cost Volume Profit Analysis

Cost Volume Profit analysis helps managers to understand the relationships between cost, volume, price, and profit by focusing on (i) selling price per unit, (ii) variable cost per unit, (iii) total fixed costs, and (iv) sales volume. Jaedicke and Robichek (1964) found that the two

typical decision categories used by business management in C-V-P analysis are: (i) decisions related to the required sales volume to attain a target profit level, and (ii) decisions related to the maximum profitable combination of products to produce and market. Management recognizes that planning and control are essential and CVP analysis is one such planning and control tool. A 2003 survey of management accounting practices found C-V-P analysis as one of the most commonly used techniques in accounting (Garg, Ghosh, Hudick, & Nowacki, 2003).

3. C-V-P Analysis and Breakeven Point under Uncertainty

3.1. Basic Breakeven Point

The initial step in C-V-P analysis is finding the breakeven point. The breakeven point occurs when total revenues equal total expenses, thus, profit is zero. At this point, total revenue (TR) minus total variable costs (TVC) and minus total fixed costs (TFC) is zero ($TR - TVC - TFC = 0$). The breakeven point can be determined in units and in dollars once the contribution margin (selling price less variable cost) or contribution margin per unit (CMU) is known, or where the contribution margin ratio [CMR], defined as the contribution margin divided by the selling price, is known. If X is the number of units sold at breakeven point, then at breakeven point, X may be determined as TFC/CMU . Dollar sales at breakeven point may be determined as TFC/CMR .

In practice, most companies target a desired net income (after tax income). To achieve this, companies target operating income (TOI) such that when the appropriate tax is taken by the tax authorities, the company will be left with the desired net income. In that case the required units to be sold is determined as $[TFC + TOI]/CMU$. The dollar sales required to achieve a targeted operating income is determined as $[TFC + TOI]/CMR$.

3.1.1. C-V-P Analysis and Cost of Capital

Kee (2007) proposed that the cost of capital can be incorporated in the C-V-P model. This requires determining the sales quantity needed to break even, as well as the sales quantity required to earn a desired profit or profit margin. Incorporating the cost of capital enables a firm's manager to evaluate alternative investment and cost structures to enhance a product's profitability. Kee (2007) concluded that "the C-V-P model based on the discounted economic income of products enables managers to compute a product's breakeven sales quantity, to measure a product's profitability over the range of its sales, and to determine the rate of change in its profitability" (p. 491).

3.1.2. Integration of Accounting and Economics in C-V-P Analysis

Xi-juan and Jun-ying (2007) proposed a stochastic Cost-Volume-Value (CVV) model with economic demand and cost functions. They indicated that the purpose of modern corporate business activities is not profit maximization (C-V-P model), but value maximization because product prices and unit costs are related to volume. They developed a suitable C-V-P model for business value creation and an actual situation under uncertainty. They concluded that the CVV model is more advanced than the traditional C-V-P model, because the breakeven point is the volume that takes all the capital costs into account.

3.2. Cost Volume Profit Assumptions and Limitations

C-V-P analysis employs a number of important assumptions. These assumptions give rise to some of the limitations of traditional C-V-P analysis. Traditional C-V-P analysis assumes that total cost can be divided into fixed costs and variable costs with respect to the level of output. During the analysis, fixed costs, unit variable cost and the unit selling price of the product remain constant. There is also the assumption that the behavior of total revenues and total costs is linear in relation to output units within the relevant range of activity. In a linear revenue situation, sales mix (the ratio of each product to total sales) and unit prices remain constant. However, selling prices and mixes can change in practice and total revenues may not be linear (Edinburg and Wolcott, 2005). Furthermore, it is often assumed that units produced are sold, which is not very realistic. Traditional CVP analysis assumes that costs, price, and volume are variables known with certainty. This may not always be the case. The random behavior of such variables creates uncertainty in breakeven analysis (Chrysafis & Papadopoulos, 2009).

3.3. Cost-Volume-Profit Analysis Under Uncertainty

The main limitation of traditional C-V-P analysis is that it does not consider uncertain features of the firm's operation. Jaedicke and Robichek (1964) presented the first CVP model under uncertainty. They allowed the variables to be random in the basic CVP model. This model is expressed by the following equation (Jaedicke & Robichek, 1964; Liao, 1975; Magee, 1975; Shih, 1979):

$$Z = Q(P - V) - F \quad (1)$$

where:

- Z : Total Profit
- Q : Unit Sales Volume
- P : Unit Selling Price
- V : Unit Variable Cost
- F : Total Fixed Costs

In this equation, the traditional CVP analysis model is appropriate for profit planning under a condition of predictable demand or when the size of demand is unlimited and unsold units can be inventoried and finally sold (Shih, 1979). There is no adjustment for risk and uncertainty. Jaedicke and Robichek (1964) stated that "the best alternative cannot be chosen without some statement of the firm's attitude toward risk" (p. 926). Therefore, understanding the amount of risk the firm is willing to accept is crucial to selecting the best alternative.

Studies introducing risk or uncertainty into C-V-P analysis have assumed normal distributions of profits (Jaedicke & Robichek, 1964; Jarrett, 1973; Kim, 1973; Johnson & Simik, 1974; Hilliard & Leitch, 1975). These studies made assumptions that demand and output were equal or unequal. Ismail and Louderback (1979) studied the implications of demand being less or greater than output by considering penalties for inadequate or excess output. They presented the optimization model for the firm by employing the general stochastic model with an uncertain

demand and known probability distribution. Other authors analyzed the random behavior of profits by using various distribution methods such as Tchebychef's theorem (Buzby, 1974), model sampling and curve fitting techniques (Liao, 1975), and lognormal distributions (Hilliard & Leitch, 1975).

3.4. Cost, Price, and Volume Properties in C-V-P Analysis

Research on the relationship between the price and costs using C-V-P analyses under uncertainty have focused on the condition of uncertainty in price and costs, and also established a relationship between price and output volume. These stochastic C-V-P analyses treated price and volume as independent variables.

Karnani (1982) presented the firm's risk-return trade off and asserted that the three elements that determine a firm's competitive strength are (1) cost position, (2) risk-return trade off, and (3) quality of information about uncertain demand and costs. In addition, he indicated that the firm's position is stronger with lower fixed costs and lower expected variable costs per unit. Also, the lower the value of the firm's risk and uncertain demand and cost, the stronger it is and the more aggressively it can compete (Karnani, 1982).

Volume of production and sales is one of the most significant causal factors in cost and profit variation. Based on the relationship between price and volume, previous studies have incorporated a demand function in a stochastic C-V-P analysis (Kottas & Lau, 1979; Shao & Jun-Ying, 2007). Chung (1993) examined the classic C-V-P problem in which volume decisions are based on stochastic information about demand. His results showed that a firm's output is an uncertain factor and management should be more aggressive in output planning.

4. Model Development of Stochastic Cost-Volume-Profit Analysis

Based on the foregoing, traditional C-V-P analysis has a major limitation; it does not simulate a world of uncertainty. In recognition of this shortcoming, accountants have attempted to use stochastic analysis to provide a better basis for profit planning. "The use of stochastic analysis in a C-V-P analysis model is a great step forward in providing more useful information for profit planning" (Liao, 1975, p. 780).

Previous studies on the stochastic C-V-P model have focused on the distribution of profit. Jaedicke and Robichek (1964) proposed a method (hereafter, the JR model) to estimate the distribution of profit in a C-V-P model. They assumed that all the variables including sales volume, unit price, unit variable cost, and fixed costs are normally and independently distributed. The basic equation of their model is Equation 1 above.

They considered two sets of assumptions. First, that Q is a normally distributed random variable, while the other variables (P , V , and F) are assumed to be deterministic. Second, total profit (Z) is also normal with mean (μ) and variance (σ) expressed as:

$$\mu_z = \mu_q(P - V) - F \quad (2)$$

$$\sigma_z^2 = \sigma_q^2(P - V)^2 \quad (3)$$

The independence assumption dictated that the mean and variance of profits will be defined by the equation:

$$E[Z] = E[Q](E[P] - E[V]) - E[F] \quad (4)$$

where

$E[Z]$: Expected value of profit;
 $E[Q]$: Expected value of sales;
 $E[P]$: Expected value of unit selling price;
 $E[V]$: Expected value of unit variable cost;
 $E[F]$: Expected value of fixed costs

and

$$V(Z) = V(Q(P - V) - V(F)) \quad (5)$$

$$\sigma_z^2 = \sigma_q^2(\sigma_p^2 + \sigma_v^2) + E^2[Q](\sigma_p^2 + \sigma_v^2) + (E[P] - E[V])^2\sigma_q^2 + \sigma_f^2 \quad (6)$$

Furthermore, the resulting profit is assumed to be normally distributed.

The issues with the JR model's estimation of the distribution of profit in the C-V-P model were identified by Hilliard and Leitch (1975) based on the following two strong assumptions. First, that sales volume, unit price, unit variable cost, and fixed costs are all assumed to be normally distributed. Secondly, that the sales price, variable costs and fixed costs are independent. The problem arising from these assumptions is that due to the normal distribution of parameters in the JR model, the results may tend to be negative when the coefficient of variation is large. Therefore, the JR model has the restriction of small coefficient of variation.

In order to overcome the above issues, Hilliard and Leitch (1975) extended the JR model based on the assumption that quantity and contribution margin ($P - V$) are lognormally distributed random variables. They also assumed that fixed costs (F) are deterministic. Their assumptions lead to a more intuitive distribution for the model's parameters, allow for dependent relationships (between sales volume, unit price, variable unit cost, and fixed costs), and permit a rigorous derivation of the distribution of the output random variable and profit. In contrast with the JR model, their model does not have the restriction of small coefficient of variation.

A follow up article by Hilliard and Leitch (1979) went beyond the approximation of just distribution of profit by developing a general methodology to obtain and utilize the exact distribution of the breakeven point for the stochastic C-V-P model. They concluded from their earlier study that, "if F and C are bivariate lognormal, the ratio of F/C is also lognormal and thus the distribution of the breakeven point is lognormal" (p. 251). The expected value of $\text{Log}[Q_b]$ is defined in terms of the original input parameters (Hilliard and Leitch, 1979):

$$\mu = \text{Log} \left[\frac{\mu_f}{\mu_c} \sqrt{\frac{CV_c^2 + 1}{CV_f^2 + 1}} \right] \quad (7)$$

where

- μ_f : expected value of fixed costs
- μ_c : expected value of contribution (C)
- σ_f : standard deviation of fixed costs
- σ_c : standard deviation of contribution
- CV_c : coefficient of variation of contribution margin which equals σ_c / μ_c
- CV_f : coefficient of variation of fixed costs which is equal to σ_f / μ_f

The variation of $\text{Log}[Q_b]$ is expressed by Hilliard and Leitch (1979) as:

$$\sigma = \text{Log} \left[\frac{(CV_f^2 + 1)(CV_c^2 + 1)}{(\rho_{fc} CV_f CV_c + 1)^2} \right] \quad (8)$$

where: ρ_{fc} is the correlation coefficient between fixed costs and contribution margin.

Hilliard and Leitch (1975) defined the following statistical relationships between price, variable cost, and fixed costs which are used here to express μ and σ in terms of these parameters.

$$\mu_c = \mu_p - \mu_v \quad (9)$$

$$\sigma_c^2 = \sigma_p^2 + \sigma_v^2 - 2\sigma_{pv} = \sigma_p^2 + \sigma_v^2 - 2\rho_{pv}\sigma_p\sigma_v \quad (10)$$

$$\rho_{fc} = (\sigma_{fp} - \sigma_{fv}) / \sigma_f \sigma_c \quad (11)$$

where

- σ_{fv} : covariance between fixed cost and variable cost;
- σ_{fp} : covariance between fixed costs and price,
- μ_p : the expected selling price,
- μ_v : the expected variable cost.

Finally, they expressed the expected breakeven quantity as follows:

$$E[Q_b] = \frac{1}{\rho_{fc} CV_f CV_c + 1} \left[\frac{\mu_f}{\mu_c} \right] (1 + CV_c^2) \quad (12)$$

Figure 1 demonstrates the effect of coefficient of variation of the contribution margin on the expected breakeven point (Hilliard and Leitch, 1979).

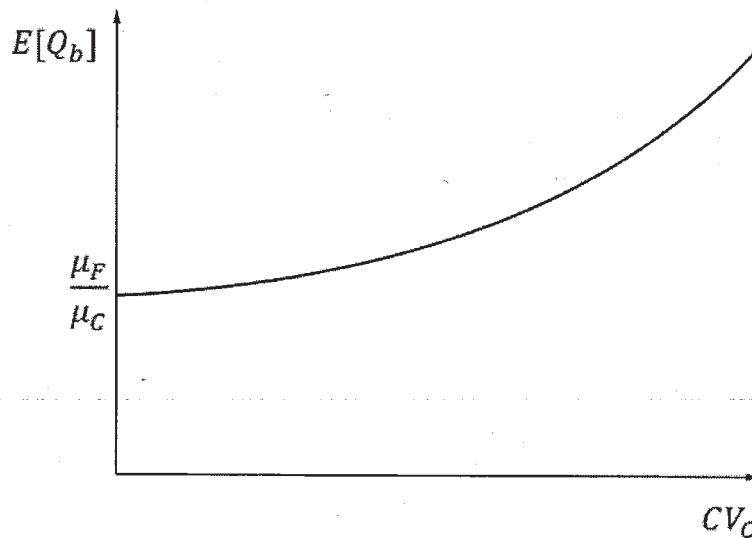


Figure 1: Effect of the coefficient of variation of the contribution margin on the expected breakeven point,

Source: J. E. Hilliard & R. A. Leitch, 1979, p. 251

In the following section the distribution of breakeven point parameters proposed by Hilliard and Leitch (1979) is examined by applying sensitivity analysis to numerical examples incorporating various selling prices and variable costs. The variations in selling price and variable cost lead to different means and standard deviations, and consequently a different contribution margin (c) is obtained. The effect of the coefficient of variation of the contribution margin on the breakeven point is analyzed.

4.1. Operationalizing CVP Under Uncertainty: Sensitivity Analyses and Examples

Sensitivity analyses of the Hilliard and Leitch model presented above are conducted to study the effect of each parameter on the expected breakeven quantity ($E[Q_b]$). In order to address the sensitivity analyses, a numerical example is presented using Hilliard and Leitch values for the basic parameters which are defined below:

$$\begin{aligned} \mu_p &= \$30, \\ \sigma_p &= 2.40, \\ \mu_v &= \$20, \\ \sigma_v &= 1.75, \\ \mu_f = F &= \$10,000 \text{ (deterministic)}, \\ \sigma_f &= 0. \end{aligned}$$

Determining the effect of the coefficient of variation of the contribution margin (CV_c) on the expected breakeven quantity ($E[Q_b]$) is the objective of this illustration and is analyzed implicitly based on the following two scenarios. The first scenario is that all the model's

parameters are constant using the values defined above except the expected selling price and expected variable cost. The second scenario is that all the model's parameters are constant using the values defined above except the variances of the selling price and the variance of variable cost. The changes in the parameters of the two scenarios lead to different coefficients of variation of the contribution margin.

As described above for the first scenario, it is assumed that the expected selling price and variable cost are unknown in the expected breakeven quantity equation. By substituting all other defined parameters in Equation 12 the equation below is obtained:

$$E(Q_b) = 10000 \left(\left(\frac{1}{\mu_p - \mu_v} \right) \cdot \left(\frac{8.822500000}{(\mu_p - \mu_v)^2} + 1 \right) \right) \quad (13)$$

The effect of the $E(Q_b)$ versus μ_p and μ_v is demonstrated by a 3D graph produced by MAPLE software in Figure 2.

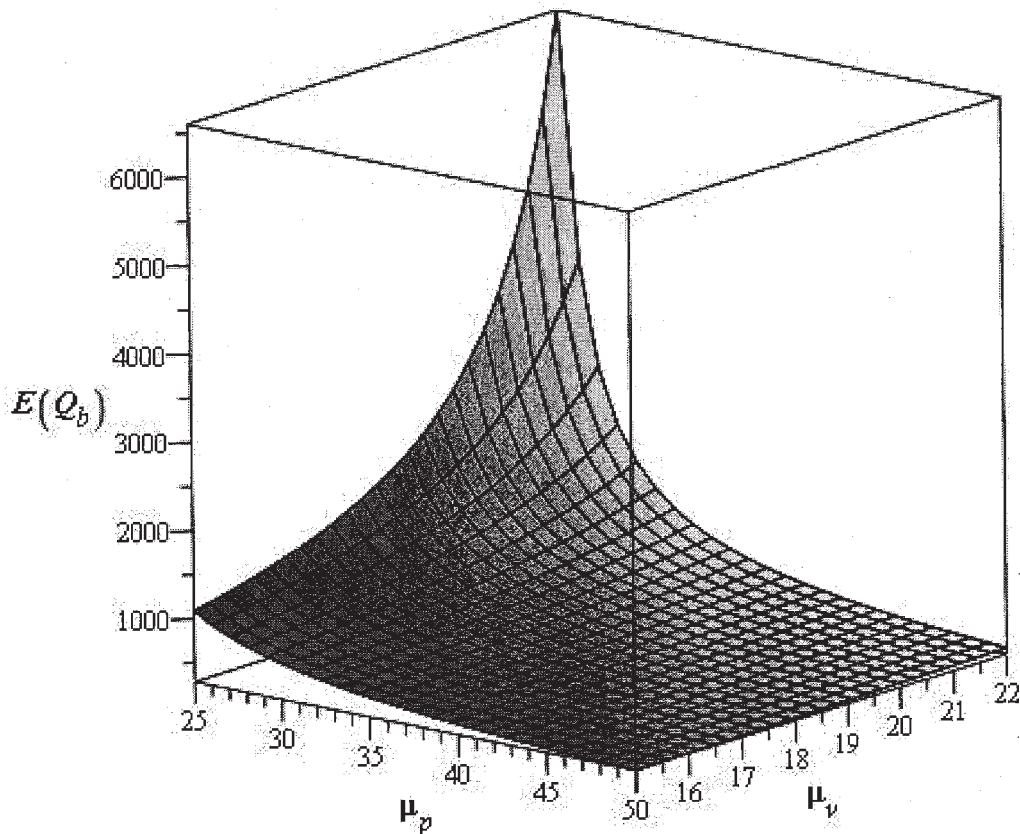


Figure 2: Expected breakeven quantity vs. expected selling price and variable cost

Analysis of the graph demonstrates that as the expected selling price decreases, near to the variable cost, the expected breakeven quantity increases rapidly. In this case, the expected contribution margin ($\mu_c = \mu_p - \mu_v$) decreases and thus the coefficient of variation of contribution margin ($CV_c = \sigma_c / \mu_c$) increases. Therefore, it can be said that as CV_c increases, $E(Q_b)$ also increases. However, the rate of change in $E(Q_b)$ is intensive when μ_p is close to μ_v . In order to better interpret this effect, Figure 3 illustrates the sequence graphs of $E(Q_b)$ as a function of μ_v based on different μ_p .

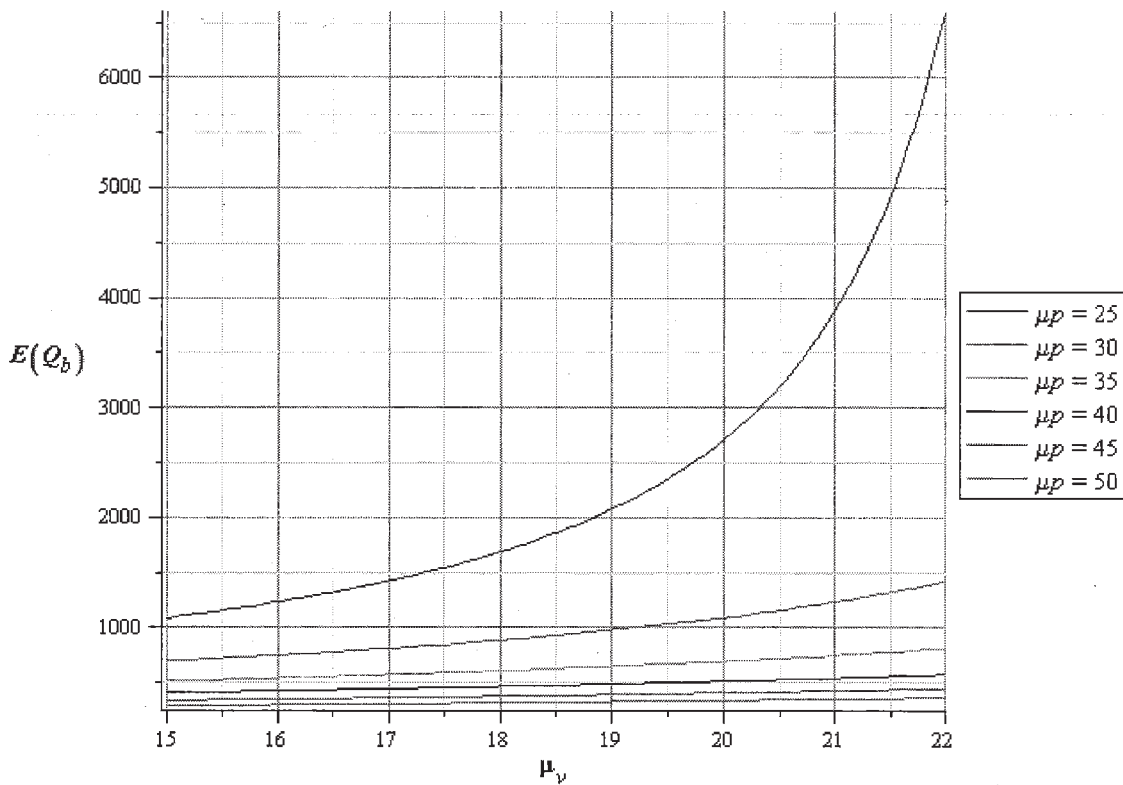


Figure 3: Effect of μ_p and μ_v on expected breakeven quantity

It can be said from Figure 3 that when the value of μ_p is far greater than the value of μ_v , differences in either μ_v or μ_p do not significantly affect the expected breakeven quantity. However, when μ_p is close to μ_v , it affects the expected breakeven quantity significantly. The following two examples clarify the above statement.

Example 1:

$$\begin{cases} \mu_p = 25 \\ \mu_v = 21 \end{cases} \rightarrow E(Q_b) = 3900, \quad \text{and} \quad \begin{cases} \mu_p = 25 \\ \mu_v = 22 \end{cases} \rightarrow E(Q_b) = 6600$$

The information in Example 1 above shows that as the μ_v increases by one unit ($\Delta\mu_v = 1$), the variation of expected breakeven quantity is equal to 2700 units ($\Delta E(Q_b) = 6600 - 3900 = 2700$).

Example 2:

$$\begin{cases} \mu_p = 30 \\ \mu_v = 21 \end{cases} \rightarrow E(Q_b) = 1250, \quad \text{and} \quad \begin{cases} \mu_p = 30 \\ \mu_v = 22 \end{cases} \rightarrow E(Q_b) = 1400$$

The information in Example 2 above shows that as the μ_v increases by one unit ($\Delta\mu_v = 1$), the variation of expected breakeven quantity is equal to 150 units ($\Delta E(Q_b) = 1400 - 1250 = 150$).

We continue to examine the effect of the coefficient of variation of contribution margin (CV_c) on expected breakeven quantity in the second scenario by changing the standard deviation of the selling price (σ_p) and the variable cost (σ_v). By substituting all other defined variables in the base numerical example, the $E(Q_b)$ is expressed as:

$$E(Q_b) = 10 \sigma_p^2 + 10 \sigma_v^2 + 100C \tag{14}$$

The problem is illustrated in Figure 4 by using MAPLE software to produce a 3D graph of $E(Q_b)$ obtained as a function of σ_p and σ_v .

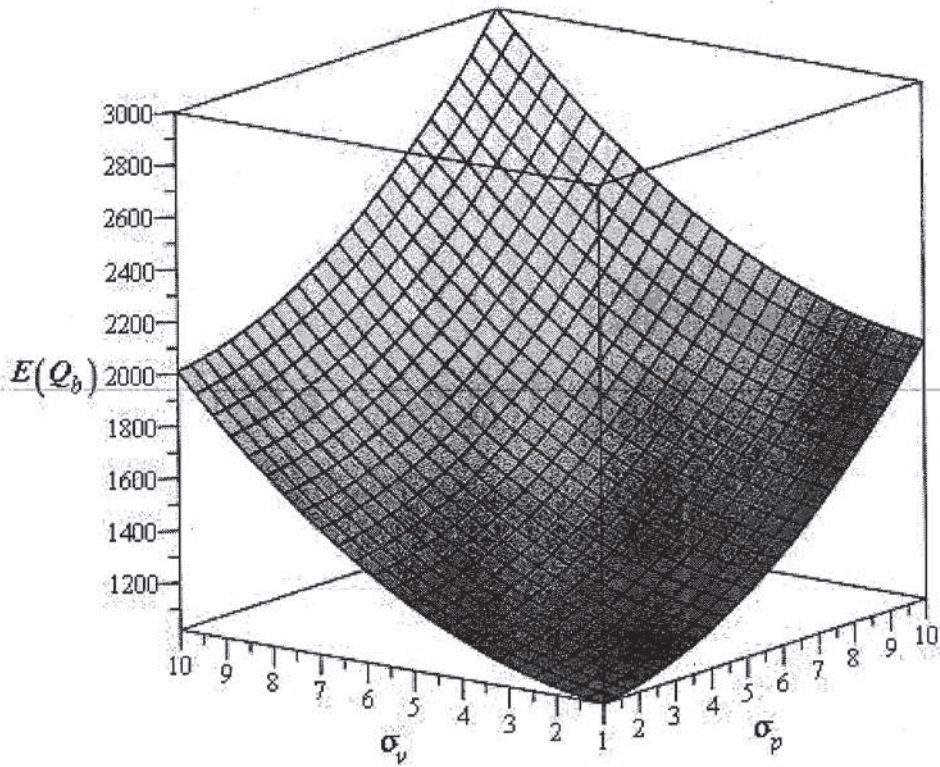


Figure 4:
Expected breakeven quantity vs. standard deviations of selling price and variable cost

It can be seen from Figure 4 that as either the standard deviation of the selling price or the variable cost increases, the expected breakeven quantity increases at the same rate. Increases in the standard deviations lead to increases in the coefficient of variation of contribution margin (CV_c), and thus expected breakeven quantity $E(Q_b)$ increases. Figure 5 illustrates the sequence graphs of $E(Q_b)$ as a function of σ_v based on different σ_p . It shows the rate of changes in $E(Q_b)$ is higher in the higher range of σ_v and σ_p .

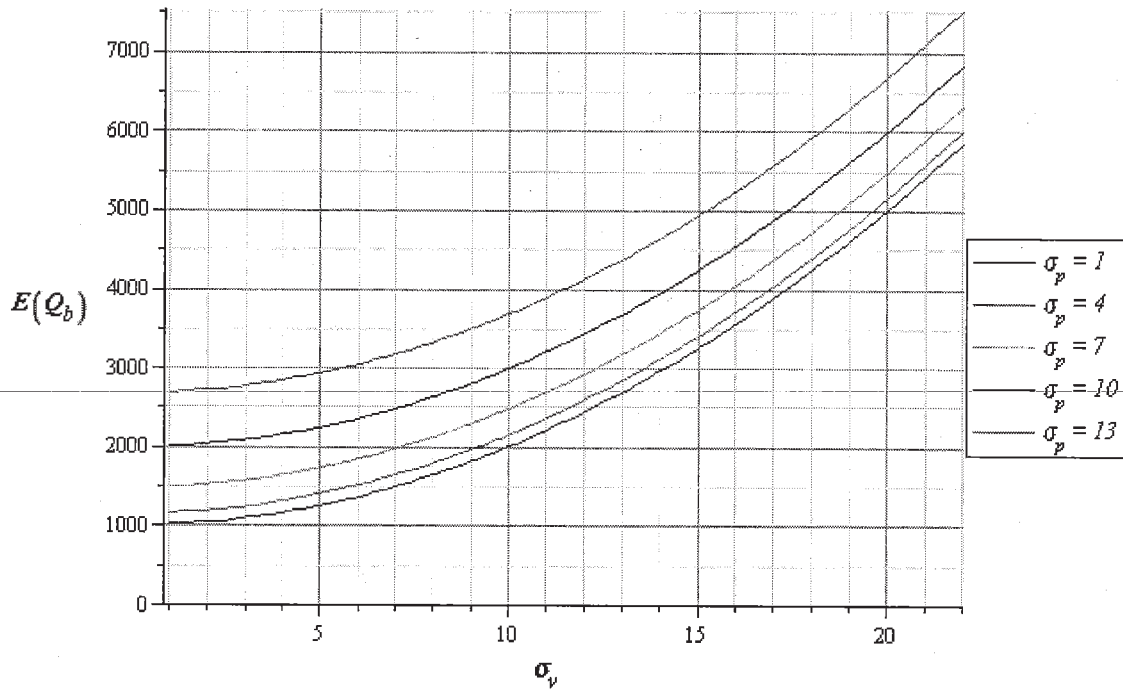


Figure 5 Effect of standard deviations of price and variable cost on expected breakeven quantity

From Figure 5, the variation of expected breakeven quantity, $\Delta E(Q_b)$, is higher when the value of either σ_p or σ_v is higher. The proof of the above statement is clarified in the following two examples.

Example 3:

$$\left\{ \begin{array}{l} \sigma_v = 15 \\ \sigma_p = 1 \end{array} \right. \rightarrow E_1(Q_b) = 3300, \quad \text{and} \quad \left\{ \begin{array}{l} \sigma_v = 15 \\ \sigma_p = 4 \end{array} \right. \rightarrow E_2(Q_b) = 3450$$

$$\Delta E(Q_b) = E_2(Q_b) - E_1(Q_b) = (3450 - 3300) = 150$$

Example 4:

$$\left\{ \begin{array}{l} \sigma_v = 15 \\ \sigma_p = 10 \end{array} \right. \rightarrow E_3(Q_b) = 4300, \quad \text{and} \quad \left\{ \begin{array}{l} \sigma_v = 15 \\ \sigma_p = 13 \end{array} \right. \rightarrow E_4(Q_b) = 4950$$

$$\Delta E(Q_b) = E_4(Q_b) - E_3(Q_b) = (4950 - 4300) = 650$$

The above results underscore the notion that $\Delta E(Q_b)$ in Example 4 is greater than the $\Delta E(Q_b)$ in Example 3.

5. Summary, Contributions and Limitations

This paper has revisited C-V-P analysis and breakeven point. Traditional C-V-P analysis has been used as one of the common accounting models in decision making and profit planning by management. This model assumes certainty in price and costs. Therefore, traditional C-V-P analysis has a major limitation; it does not usually incorporate uncertainty. For this reason, stochastic C-V-P analysis is used to provide a better basis for profit planning.

We have analyzed a stochastic C-V-P model with a distribution of breakeven quantity. The analyses concentrated on the Hilliard and Leitch (1979) study to obtain the exact distribution of the breakeven quantity for the stochastic C-V-P model by considering lognormal distribution of parameters (F , C , Q_b). Furthermore, the effect of the coefficient of variation of the contribution margin (CV_c) on the expected breakeven point quantity ($E[Q_b]$) was determined and analyzed explicitly based on two cases:

- Expected breakeven quantity versus expected selling price and expected variable cost.
- Expected breakeven quantity versus the standard deviation of selling price and the standard deviation of the variable cost.

In the first case all the variables were assumed constant except the expected selling price and expected variable cost. It showed that when expected selling price μ_p decreases, the expected breakeven quantity $E(Q_b)$ increases in an exponential function.

In the second case all the variables were assumed constant except the standard deviations of the selling price and the variable cost. Results showed that as the standard deviation of selling price or variable cost increases, the coefficient of variation of contribution margin (CV_c) also increases. Thus the expected breakeven quantity $E(Q_b)$ increases at the same rate.

5.1 Contributions, Recommendations, Limitations

The first major contribution of this work is the use of the Maple Software to operationalize the theoretical models of CVP under uncertainty so that they can be used in practice. Most managers will agree that the certainty and linearity assumptions of the traditional CVP model are sometimes not realistic. However, ways to incorporate uncertainty have largely remained academic and theoretical, thus preventing their practical application. Our graphical demonstrations bring them to the floor level by providing visual demonstrations of the effects of changes in the behavior of CVP variables. Given this possibility, it is recommended that businesses that use CVP analysis extensively, especially in competitive industries and markets, should incorporate more uncertainty into their planning assumptions. CVP analysis and planning need not be limited to certainty assumptions.

The major limitation of this study is that it did not use real life business data in demonstrating the applications of the various examples presented. This can be the subject of further studies. Secondly, the normal or lognormal distribution of the breakeven point presented in this study may not be assertable. The current state of research on this limitation is inadequate and requires further research.

REFERENCES

- Brockett, P. L., Charnes, A., Cooper, W. W., & Shin, H. C. (1984). A Chance Constrained Programming Approach to Cost-Volume-Profit Analysis. *Accounting Review* , 59(3), 474-487.
- Brockett, P. L., Charnes, A., Cooper, W. W., & Shin, H. C. (1987). Cost-Volume-Utility Analysis with Partial Stochastic Information. *Quarterly Review of Economics and Business* , 27(3), 70-90.
- Chrysafis, K. A., & Papadopoulos, B. K. (2009). Cost-Volume-Profit Analysis Under Uncertainty: a Model with Fuzzy Estimators Based on Confidence Intervals. *International Journal of Production Research* , 47 (21), 5977-5999.
- Chung, K. H. (1993). Cost-Volume-Profit Analysis Under Uncertainty When the Firm has Production Flexibility. *Journal of Business, Finance and Accounting* , 20(4), 583-591.
- Chung, K. H. (1990, 36 November). Output Decision under Demand Uncertainty with Stochastic Production Function: A Contingent. *Management Science* , 1311-1328.
- Edinburg, L. G., & Wolcott, S. K. (2005). *Cost Management: Measuring, Monitoring, and Motivating Performance*. New York: John Wiley & Sons.
- Garg, A., Ghosh, D., Hudick, J., & Nowacki, C. (2003). Roles and Practices in Management Accounting Today. *Strategic Finance* , 85(1), 30-35.
- Hilliard, J. E., & Leitch, R. A. (1979). A Stochastic Analysis of Breakeven Points. *AIIE Transactions* , 11 (3), 250-253.
- Hilliard, J. E., & Leitch, R. A. (1975). Cost-Volume-Profit Analysis under Uncertainty: A Log-Normal Approach. *The Accounting Review* , 39, 917-926.
- Ismail, B. E., & Louderback, J. G. (1979). Optimizing and Satisficing in Stochastic Cost-Volume-Profit Analysis. *Desicion Sciences* , 10, 205-217.
- Jaedicke, R. K., & Robichek, A. A. (1964). Cost-Volume-Profit Analysis under Conditions of Uncertainty . *The Accounting Review* , 39, 917-926.
- Jarrett, J. E. (1973). An Approach to Cost-Volume-Profit Analysis under Uncertainty. *Decision Sciences* , 4(3), 405-420.
- Johnson, G. L., & Simik, S. S. (1974). The Use of Probability Inequalities in Multiproduct C-V-P Analysis under Uncertainty. *Journal of Accounting Research* , 12(1), 67-79.
- Karnani, A. (1982). Stochastic Cost-Volume-Profit Analysis in a Competitive Oligopoly. *Desicion Sciences* , 14, 187-193.
- Kee, R. (2007). Cot-Volume-Profit Analysis Incorporating the Cost of Capital. *Journal of Managerial Issues* , 19 (4), 478-493.
- Kim, C. (1973). A Stochastic Cost Volume Profit Analysis. *Decision Sciences* , 4(3), 329-342.

- Lau, A. H.-L., & Lau, H.-L. (1987). CVP Analysis with Stochastic Price-Demand Functions and Shortage-Surplus Costs. *Contemporary Accounting Research* , 4(1), 194–209.
- Liao, M. (1975). Model Sampling: A Stochastic Cost-Volume-Profit Analysis. *The Accounting Review* , 50(4), 780-790.
- Magee, R. P. (1975). Cost-Volume-Profit Analysis, Uncertainty and Capital market Equilibrium. *Journal of Accounting Research* , 257-266.
- National Association of Cost Accountants. (June 15 1914). The Variation of Costs with Volume. *N.A.C.A. Bulletin* , 30(16), 1215.
- Shih, W. (1979). A General Decision Model for Cost-Volume-Profit Analysis under Uncertainty. *The Accounting Review* , LIV (4), 687-706.
- Xi-juan, S., & Jun-ying, F. (2007). Stochastic Cost-volume-value Model with Economic Demand and Cost Function. *IEEE* , 3927-3930.
- Yunker, J. A. (2001). Stochastic CVP Analysis with Economic Demand and Cost Function. *Review of Quantitative Finance and Accounting* , 17, 127–149.
- Yunker, J. A., & Yunker, P. J. (1982). Cost-Volume-Profit Analysis Under Uncertainty: An Integration of Economic and Accounting Concepts. *Journal of Economics and Business* , 34(1), 21-37.

ASB 2013
Antigonish, Nova Scotia

Lawrence T. Corrigan (student)
Sobey School of Business
Saint Mary's University

THE QUANTITATIVE-QUALITATIVE DEBATE: ADD TEMPORALITY AND STIR

The quantitative-qualitative debate in Management and Organizational Studies has been of interest to scholars over the years. However, temporality seems to be absent from the discussion. This paper reviews the debate and its epistemological and methodological foundations. The paper supports the idea that research *methods* are different in degree and different in kind but that, fundamentally, *research* is neither quantitative nor qualitative. Despite the overwhelming incidence of studies that use stand-alone quantitative research methods, the paper concludes that it is doubtful such studies can effectively deal with the temporality of social variables unless qualitative enquiry is also employed.

Introduction

'Time has no being since the future is not yet, the past is no longer, and the present does not remain.' (Ricoeur, 1984, p. 7)

The 2013 Atlantic Schools of Business Conference (*Business in Time*) calls for papers that reflect upon the discourse of time. The theme of temporality resonates strongly with social scientists who are concerned with the validity of their production. For quantitative researchers, 'history' is an epistemological threat that events occurring concurrently with treatment could cause the observed effect (Shadish, Cook, & Campbell, 2001). Qualitative researchers tend to see time and history in quite a different light, viewing these as situated, for example purporting that knowledge is "influenced by its temporality" (Weatherbee & Durepos, 2010, p. 18). The dominance of the quantitative paradigm probably means that qualitative researchers will find it difficult to "get by without some (sometimes even quite substantial) knowledge of quantitative methods and methodological standards, whereas in several disciplines there is no immediate need for quantitative researchers to 'bother' much with qualitative methods" (Fielding & Schreier, 2001, p. 2). Across research traditions, respondents cannot be totally isolated from outside influences, resulting in knowledge production processes that have to make validity assumptions about ways of knowing. One important cluster of assumptions relate to temporality.

As a recurring theme in the academic literature, some scholars have claimed that research is fundamentally neither quantitative nor qualitative, or at least that these traditions are not incompatible with each other (e.g., Howe, 1988; Kritzer, 1996; Madill & Gough, 2008; Nash, 2002). The rhetoric of the quantitative-qualitative debate covers a wide gap. Some scholars suggest that all research methods share fundamental similarities while others insist that various research methods are worlds apart. What seems to be missing in the debates is the notion of time. Organizational studies often conceptualize time and history as a knowledge-building process

(although the conceptualization may not be specifically recognized) which passes through a linear series of events (Aminzade, 1992; Griffin, 1992; Munslow, 2010). There may also be an implied assumption “that meaning is carried through time” (Cunliffe, Luhman, & Boje, 2004, p. 261). In Management and Organizational Studies there is an increasing call for a historical perspective (e.g., Clark & Rowlinson, 2004). Temporality is a ubiquitous ingredient, often taken as known even though it can be seen as either an objective or subjective concept. Temporality should be explicitly considered as researchers try to make sense of their methods and results.

In this paper, I engage with the quantitative-qualitative debate by adding the notion of temporality which has been nearly absent from the discussion to date. Most historians assign particular importance to the explanatory role of time (Griffin, 1992) and organizational researchers (along with historical sociologists) have censured their discipline for tending to ignore time and history when studying the social (Aminzade, 1992). They may do this “either by studying the correlates of outcomes rather than the character of temporally connected events or by treating events as surface manifestations of large-scale and long-term processes of change” (Aminzade, 1992, p. 456). It seems in order, then, to ask the following question:

How does consideration of temporality shed light on the interdependence of quantitative and qualitative research?

This paper argues that, fundamentally, *research* is neither quantitative nor qualitative; however, research *methods* can be shown to be rather different in degree and different in kind. Although the paper is predicated on the notion that quantitative and qualitative studies are more similar than different, a case can be made that qualitative considerations are always needed to take account of temporality. Therefore, despite the overwhelming incidence of studies that use stand-alone quantitative research methods, it is doubtful that such studies effectively deal with the temporality of social variables without adding qualitative enquiry. This paper proceeds as follows: first, I introduce the quantitative-qualitative debate; second, the similarities and differences of these research traditions are discussed; third, philosophical foundations are explored from the point of view of temporality; finally, I comment on the methodological strand of the debate and conclude that adding temporality to the quantitative-qualitative debate suggests that quantitative research cannot go-it-alone in explaining organizational society.

The quantitative-qualitative debate

At the outset, the following clarification of the two key terms is offered (although these brief definitions are by no means unproblematic). By *quantitative* I refer to the mainstream search for ‘explanation’ based on statistical modes of processing data (especially numeric data) and generally attempting to follow practices of the natural sciences. By *qualitative* I refer to methodological approaches that rely on nonquantitative (or nonstatistical) modes of processing data in a search for ‘understanding’ that is based on social construction of knowledge. Above I speak of ‘debate’ in the singular but hasten to acknowledge that for at least the past 50 years the academic literature has debated various contests between quantitative and qualitative issues. For example there has been extensive debate (in a variety of fora) over the relative merits of different modes of research. Aldrich (1988, p. 19) refers to the antagonists in these disputes as ‘paradigm

warriors'. Even within categories there is debate. For example, Kidder and Fine (1987) have ascribed two meanings to qualitative: the big Q and the small q. Qualitative work with a big Q consists of an evolving set of questions without a structured design. Qualitative work with a small q includes open-ended questions in a survey that has structure. There is also the compatibility debate (see Fielding & Schreier, 2001; Howe, 1988) that discusses whether quantitative and qualitative research is similar in temperament at either the level of epistemology or practice. This is the conversation that I wish to enter, bringing along with me the notion of temporality.

Research is considered to be a craft, requiring soft skills such as creativity and adept use of language. In qualitative research the “messy, random and creative elements” (P. Prasad, 2005, p. 6) give many of its traditions a distinctive artistic character. In quantitative research there is also tremendous scope for crafting, including the selection and implementation of analytic tools such as factor analysis, multidimensional scaling, multivariate analysis, etc., that provide “all sorts of opportunities for ‘playing’ with the data” (Kritzer, 1996, p. 24). Sometimes this creatively reveals unexpected patterns over time. Donna Haraway (1994), a poststructuralist writer, compares this possibility to the game of cat’s cradle where research colleagues can build-up knotted string patterns and pass along some unexpected results. This can be done after playing with the data over different time(s) and space(s).

Quantitative and qualitative – these terms are sometimes used in odd contexts and, as indicated above, some scholars see the two as not very different from each other. It is instructive that chemists (ultra functionalists) use the term “qualitative analysis” when determining what a chemical is made of – really a search for the inner constitution of the chemical. Time is often specifically considered in chemistry experiments. For example, “stochastic time evolution” is used to study the behavior of chemical molecules; however, in looking at the reactions “it is not at all obvious just how we should go about using [hypotheses] to calculate this temporal behavior” (Rotermund, Engel, Kordesch, & Ertl, 1990, p. 2343). Time is also problematic in Management and Organizational Studies where it is often reduced to the status of a bystander (although perhaps not an innocent bystander), for example as one of many independent variables in a multivariate regression analysis. The idea ‘that quantitative and qualitative research is essentially the same thing’ can be practically shown by looking at some typical research questions. Surely, research questions are at the core of academic activity. For example:

Qualitative researchers in the area of Critical Management Studies address questions such as:

- What is the source of organizational power?
- How does management control the workplace?
- How do employees resist organizational change?

Quantitative researchers in the area of Organizational Behaviour ask questions such as:

- What causes firm financial performance?
- What are the moderating effects of gender on a particular outcome?
- What are the relationships among variables?

These research questions would be dressed-up differently by the two types of researchers but the basic nature of such questions is shared among research traditions, i.e., the quest for knowledge about individuals and organizations. Here is the important point; the above research questions could be easily swapped among researchers of different research traditions and still be of academic interest. There is an excellent story of the principle of the drunkard's search (Howe, 1988); a drunkard was searching under a street lamp for his house key, which he had dropped some distance away. Asked why he didn't look where he had dropped it, he replied, "It's lighter here!" Those who insist that quantitative and qualitative research is incompatible, like the drunkard's search, call more attention to how to look for something, rather than focusing on the more important questions of where the evidence might be and why it is being looked for.

Research questions can be framed as differences in kind or differences in degree. This provokes an important discussion. However, I don't see these differences as mutually exclusive. People tend to put too much emphasis on polar extremes; for example, seeing categorical measurement as strictly qualitative and continuous measurement as only quantitative. In organizational behaviour research we often hear of the dichotomy between introverts and extraverts, but there probably is no pure form of these when temporality is considered. Susan Cain gave a lecture (TED Talks, 2012) on introverts and extroverts, saying that any differences fall along a continuum and movement along the continuum is possible. When temporality is considered, after a while it may become apparent that a big enough shift in the 'difference of degree' starts to look more and more like a 'difference in kind'.

Similarities and differences

This paper is predicated on the notion that quantitative and qualitative studies are more similar than different. However, there would be no shortage of opponents to this view, including on methodological grounds. Comparisons of methods are often presented as conflicting approaches, and the conflict has fostered paradigm segregation among academic communities (discussed in the next section). Practical matters also contribute to the dichotomy. Computer software is usually packaged to appeal to separate interests. For example, *atlas.ti* is designed for qualitative analysis (and marketed that way) and *SPSS* for quantitative analysis. The quantitative-qualitative contest is often characterized as numbers versus words, respectively. This dichotomy is sometimes presupposed, as in many statistics textbooks. For example, Zikmund et al. (2010, p. 134) state that the quantitative *versus* qualitative debate [emphasis added] is "largely a superfluous argument in either direction" but then subsequent text material is partitioned into those two categories. Discourse in the literature sometimes exhibits a combative tone; quantitative researchers who claim to value statistical significance over "navel gazing" and qualitative researchers who prefer situated knowledge to "number crunching" (Kidder & Fine, 1987, p. 57). Such comments not only declare preferences but also a distrust of the work of others. Nash (2002, p. 409) laments such "ruptures... for the terms 'qualitative' and 'quantitative' are not semantic antonyms, and continues to inhibit principled attempts to develop integrated 'numbers and narratives' models". Narrative is not about numbers versus words but is a story of social phenomena as temporally ordered and the story is subject to retelling (Abbott, 1992).

Those who profess that various research modes share fundamental similarities would likely admit that the look and feel of different research traditions is sometimes worlds apart. The following excerpts from actual research results from management journals illustrate the point:

- #1a: *“Organizational commitment remains significant when controlling for trust, consistent with the hypothesis, that trust only partially mediates the relationship between perceived organizational support and organizational commitment.”*
- (Whitener, 2001, p. 527)
- #1b: *“Table 2 presents means, standard deviations, and correlations. Turnover averaged 18.36% per year, and the logarithm of the productivity averaged 12.05, or annual sales of \$171,099 per employee. The mean q was .46 and the gross rate of return was 5.10%.”*
- (Huselid, 1995, p. 654)

- #2a: *“Implications for the management theorist as they pertain to a more fragile tone for “doing history” include the acknowledgement of history as multiple socially constructed interpretations of the past.”*
- (Durepos, Helms Mills, & Mills, 2008, p. 116)
- #2b: *“It means both building and destroying machines, identities, categories, relationships, space stories. Though both are bound in the spiral dance, I would rather be a cyborg than a goddess.”*
- (Haraway, 2006, p. 147)

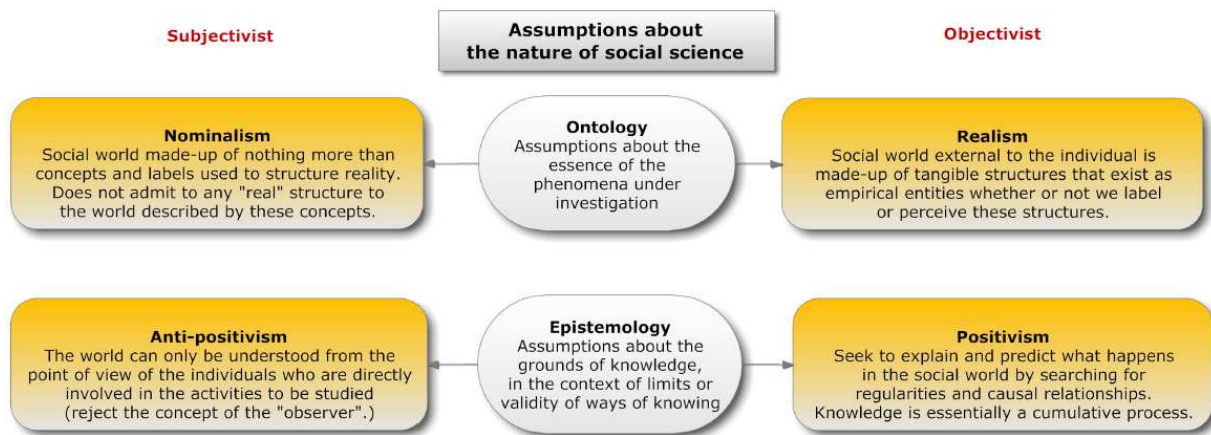
It may be difficult to articulate all the subtle differences between quantitative and qualitative research, although (as illustrated in the above “look and feel” examples) we can easily recognize the difference when we see it. The positivist results reported in #1a and 1b have a different tone than the interpretivist results in #2a and 2b. Even in these examples, both ‘kinds’ of researchers construct arguments based on the data that were collected and are aware that alternative explanations may exist.

The central role of interpretation in qualitative research is well documented. But most quantitative data in social science are also constructed through subtle interpretation (Kritzer, 1996). For example, the administering of questionnaires to human respondents builds-in interpretation by design. The researcher makes judgment calls in the process of writing survey questions, constructing scales, applying codes, and deciding whether all the data will count as an acceptable response. When and how to delete data is open to alternative approaches such as listwise deletion or pairwise deletion (or no deletion!) From a historical perspective, the important point is that much research data is considered to be irrelevant or even harmful to the analysis. In qualitative research using time as a historical construct, Burke (2012) refers to the possibility of losing knowledges and what Burke (p. 147) calls “collective forgetting”. This raises the question as to how possible is it to demonstrate what counts as knowledge when versions of ‘the truth’ change over time. The processes of knowledge-losing also include destroying and hiding

knowledges, for example the practice of public organizations to hold private “in camera” meetings. Burke (2012) gives a historical summary of political hiding in the economic domain, for example the routine classifying of ‘official secrets’ in 50-year dormancy rules. Burke also discusses issues of censorship throughout a 250 year period and the modern use of ‘firewalls’ to block access to computer networks, for the sake of security. Knowledge loss is certain to remain a temporal consideration in both quantitative and qualitative research.

Philosophical foundations and paradigms

At the centre of the quantitative-qualitative debate is a philosophical discussion. Unfortunately much of the attention drawn to methodological divergence derives from a tendency for philosophical and technical issues to be treated simultaneously and occasionally to be confused (Bryman, 1984). Assumptions about the nature of social science are elaborated in a seminal (albeit somewhat outdated) book by Burrell & Morgan (1979, hereafter B&M). With my apologies to the authors for crudely summarizing their richly decorated arguments into one-sentence strands of thought, the following chart outlines alternative approaches to assumptions about the essence of the phenomena under investigation (ontology) and assumptions about the grounds of knowledge (epistemology).



B&M articulate a subjectivist-objectivist dimension that makes evident the assumptions that qualitative and quantitative researchers use. Ontologically, qualitative research employs interpretive sociology and implies a nominalist stance, understanding reality as socially-constructed through words, behaviours, texts and other symbolic resources. Such a constructivist attitude suggests that ‘the world’ is real only to the extent that we make it so. Different researchers may constitute different social worlds since there can be many ‘realities’. Poststructuralism (e.g., Foucault, 1982) deepens the exploration of reality by emphasizing discontinuity, complexity and randomness of events. On the other hand, realism implies a world

'out there' which can be known by social scientists. Social phenomena are treated as concrete and external to the individual. Realist ontology has been (and still is) the dominant framework for the study of organizations. Researchers within this functionalist sociology are concerned to provide explanations for the status quo and social order. The approach is pragmatic and problem oriented. This brief explanation oversimplifies the philosophical elements of ontology (and ignores B&M's second dimension – the sociology of regulation and radical change). The basic difference for purposes of this paper is that subjectivist researchers tend to see reality as being different things for different people; on the other hand, once objectivist researchers find reality, they should all find the same thing. In terms of epistemology, qualitative research tends to be anti-positivist, and quantitative research is closely tied to positivism. This has important implications for how research is actually carried out. It is usual for qualitative researchers to attempt to see the social world from the point of view of the actors, and reflexively seeing themselves as implicated in their own research. Quantitative researchers tend to inquire from the outside, looking for causal relationships and generalizability.

The subjective-objective distinction also involves aspects of temporality including pace, duration, cycle, and trajectory (Aminzade, 1992). Functionalists usually measure objective time, based on fixed points of view and calculated as a factor of the rotation of the earth on its axis. However, "sociologists of time insist on the subjective features of time, its nonlinearity and social malleability" (Aminzade, 1992, p. 470). Quantitative studies have mostly focused on the materiality of time, and how it can objectively structure social action. However, "in the realm of psychological experience, quantifying units of time is a considerably clumsier operation. It is this usually imprecise psychological clock, as opposed to the time on one's watch, that creates the perception of duration that people experience" (Cunliffe et al., 2004, p. 266). These objective and subjective positions are reflected in research viewpoints. How we conceive of time has a major influence upon our ideas of process. Objectivist researchers usually seek generalizability, i.e., they do not limit their inferences to the contexts of time and space from which their data were drawn. Indeed, they extensively document how their samples are representative of a greater population, even if the research occurs at only one point in time and/or in one place (Griffin, 1992). The mainstream concern is to interrogate variables with statistical tools to show causal outcomes. Often the only temporal criterion for inferring a cause and effect relationship is that the presumed cause must precede the effect. However, one important criticism is that quasi-experiments merely provide a black box view of causality, especially where the actors in a particular treatment may vary in their experience of the temporal flow of events.

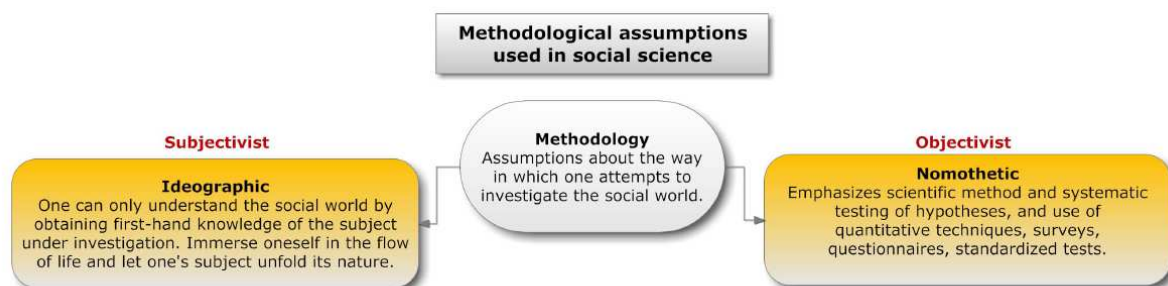
B&M (1979) argue that all social theorists are situated in a particular paradigm which defines (a limited range of) intellectual territory. There is room for variations within each paradigm although the four paradigms described by B&M are claimed to be mutually exclusive. This view is not accepted by everyone. In practice, researchers mix quantitative and qualitative data and "so-called quantitative studies are pregnant with (ontologically) qualitative concepts" (Howe, 1988, p. 15). Logic informs all academic research, whether it is hypothetico-deduction associated with many quantitative studies or the sensemaking associated with qualitative work. Sometimes the appeal to logic takes the form normally presented by the positivist view of natural science, with its insistence on hypotheses, verifiability, generalizability and other trappings of the 'scientific method'. As indicated above, positivistic epistemology has gained prominence, and

other research methodologies that fail to comply with such a standard are dismissed as unscientific. The paradigm disputes take a toll on qualitative researchers who sense a practical need to be accepted by the positivist mainstream (for example, to achieve tenure or promotion in university appointment processes). Prasad (2005, p. 4) refers to this as “positivist anxiety”. Another useful term is “qualitative positivism” (A. Prasad & Prasad, 2002, p.6) which describes research employing qualitative methods such as interviews, observation, participant observation, ethnography, etc. within positivist assumptions about the nature of social reality.

What counts as evidence is partially determined by the ontological and epistemological context in which management ideas develop. But research divisions do not split only along the qualitative-quantitative divide. There is a large amount of competition within paradigms, for example B&M segments the interpretive paradigm into four parts: solipsism, phenomenology, phenomenological sociology and hermeneutics. It seems that the various proponents do not speak directly to those favouring another paradigm, except to attempt to discredit. The framework of B&M does not recognize that well-established traditions such as phenomenology and feminism transcend disciplinary boundaries. Also, more recent intellectual schools of thought such as poststructuralism (e.g., actor-network theory) were not considered in the B&M framework. The paradigm conflict that exists in the academic literature has much of its basis in the similarities/differences of methods typically cast as quantitative or qualitative. It is not likely that positivist organizational theory will heavily integrate subjectivist views within the ‘normal science’ stream. The meta-theoretical assumptions that underpin different paradigms are complex (Moran-Ellis et al., 2006) but there is a call in social science for greater use of multiple study methods to increase the credibility and validity of reported results. This opens-up the question of how researchers deal with the qualitative-quantitative divide in their data sources and methods, which I now turn to in the next section.

Methodology as a strand of the debate

B&M (1979) propose that a further strand of debate centres on the methodological assumptions used in social science. B&M use the subjectivist-objectivist dichotomy for this line of argument. The main points of discussion, as they relate to the quantitative-qualitative divide, are outlined in the following chart (again, crudely summarized from the work of B&M):



If qualitative researchers in the subjectivist mode make a commitment to the epistemological principles of social science research discussed earlier in this paper, then research methods must facilitate an inside view that considers time. For example, unstructured interviewing is a technique that enables first-hand knowledge and immersion in the flow of organizational activities, since the participants themselves help structure the 'data' by talking openly. In this mode, they do not choose from predetermined alternative answers or Likert scales. The researcher has to engage with histories (and other stories) and be concerned not only with what is told, but also the temporality and assembly of what is told. History in Management and Organizational Studies usually assumes an objectivist stance – a knowable past reality. Such methodology looks for "truth [which is] achieved through referential correspondence between the 'facts' and the past – where past and history are synonymous. However, the past and referent history are of ontologically different status" (Weatherbee, 2012, p. 205). The process of data analysis, whether the data are quantitative or qualitative, involves an interaction between the analyst and the data. Choices are made at every step of the research process, from the choice of research approach (e.g., taking a deductive approach), the formulation of the research question (e.g., assuming reality is external and objective), and the interpretation of the data (e.g., using SPSS to calculate the statistical significance of relationships). This raises the question of personal values of the researcher and understanding of potential moderating effects of time.

This paper advocates the use of qualitative methods to surface knowledge about temporal aspects of organizing. The idea is illustrated by briefly outlining a few frequently used qualitative research methods; ethnography, structured observation and text analysis:

(1) Ethnography involves immersion in the research setting for an extended period of time. It enables engagement with temporal concepts such as pace and duration. Pace refers to the number of events per segment of time, and duration is elapsed time for a given occurrence or sequence (Aminzade, 1992). Other temporal concepts such as trajectory and cycles (Aminzade, 1992, p. 459) "suggest a more qualitative understanding of time, involving pattern or direction". Cycles relate to repetitive occurrences that define a sequence, and trajectory refers to accumulation and possible trending. Taking account of organizational events requires careful study of history, which may be impossible to fathom without qualitative inquiry. Failure to study history diminishes our ability to understand organizational changes. We need an inside view of how fast, as well as how far, people are prepared to puncture social boundaries (Aminzade, 1992).

(2) Structured observation employs rules for observing behaviour and the recording of field notes. The word 'observation' may seem uncomplicated but as pointed-out by Burke (2012, p. 35) it is not "just another word for 'looking' - and the practice may seem timeless, whether we think of travellers, healers or stargazers... I should like to draw attention to the historicity of observation, including not only the increasingly rapid rise of aids to observation but also a growing awareness of the problems raised by the practice." In quantitative research, both the 'structure' and the 'observation' in structured observation would likely be seen as problematic, i.e., interference by the researcher. However, this qualitative method sheds considerable light on situated temporality.

(3) Text analysis is the study of language, such as discourse, conversation, written work and a host of other phenomena treated as text (e.g., videos, music, advertising, etc.). History text reveals

the constraints and effects of narration with respect to “the timing of the history text through the narrative distortion of duration.... [which is] most apparent when the historian speeds up or slows down its passage within the narrative in relation to the events told” (Munslow, 2010, p. 163). As researchers narrate their findings there is an unavoidable temporal distortion. There is a tendency within the quantitative tradition to completely ignore this, especially since analysis of text may be characterized as unscientific.

As indicated above, research methods share some fundamental similarities. Recognition of this helps us to remain hopeful for a greater coalescence. Currently, qualitative methods are sometimes used in quantitative research, albeit with different focus. On balance, though, it should be noted that anti-positivist (or postpositivist) researchers primarily engage with qualitative methods, and positivists emphasize quantification in the collection and analysis of data. However, the words versus numbers comparison is inadequate to explain how types of research are carried out. A case can be made that qualitative and quantitative research has a mutually beneficial relationship – for example, qualitative work can assist quantitative research by building a theoretical base, helping to explain survey responses, interpreting statistical results, and by providing case studies. However, sometimes security issues requiring anonymity (from the point of view of the organization being studied) may result in history being neglected or disguised. Anonymous respondents are often statistically claimed to be a general case and ‘timeless’ rather than actors in “some particular historically situated organizational and occupational world existing at the time of the study” (Clark & Rowlinson, 2004, p. 345). Positivist researchers need not rely only upon quantitative approaches to learn the whole story. For example, interviews can be carried-out which are then analyzed with statistical software. Qualitative studies can also be used as exploratory work to form the basis for the next quantitative phase of the research project. Also, as proposed in this paper, qualitative work can help the quantitative researcher to understand the temporality inherent in the research platform.

The discussion in this paper leads to a consideration of how quantitative and qualitative methods might benefit from each other’s strengths. The literature has a robust discussion of this idea under terms like “mixed methods” and “triangulation”. One indication of the importance of this idea is the recent (2007) establishment of the *Journal of Mixed Methods Research* which is devoted to the subject. Bryman et al. (2011) reconcile the ontological and epistemological assumptions discussed earlier, with the broader concept of ‘world views’ that are either positivist or postpositivist. This reconciliation promotes the opinion that there is no inherent incompatibility between methods and opens-up several lines of possibility such as the following three aspects:

- (1) Triangulation – this can take any of several approaches (Fielding & Schreier, 2001). Triangulation can mean the mutual validation of results obtained by different methods (cross-checking) or triangulation as a means toward obtaining a more complete analysis, and triangulation “in its original trigonometrical sense, indicating that a combination of methods is necessary in order to gain any (not necessarily a fuller) picture of the relevant phenomenon at all” (Fielding & Schreier, 2001, p. 5).
- (2) Quantitative research that facilitates qualitative research – for example, the results of statistical analysis may help screen large numbers of people in an area of research interest

to determine who should be qualitatively interviewed. The most important point, here, is that items under study (for example, various demographic issues, historical issues, natural mood states, peer effects, timing issues, etc.) transcend any particular research method.

- (3) Qualitative research that facilitates quantitative research – for example, this can include an open-ended approach to developing hypothesis, or providing deep context to help design questionnaires and determine an approach to administering the questionnaires (for example, subjectively understanding what an experiment ‘means’ to the participant at various point in time, or how historical writings influence the opinions held by research informants).

Shadish, Cook & Campbell (2001) state that qualitative traditions can uncover site-specific threats to the all-important requirement of internal validity. Threats to internal validity are reasons why inferences that the relationship between variables is causal may be incorrect. Identifying validity threats is always context specific and includes several time-related threats. For example, *ambiguous temporal precedence* is the lack of clarity about which variable occurred first. This may result in confusion about which variable is the cause and which is the effect. Also, we need to be aware that some causation is bidirectional, for example incarceration that causes criminal behaviour that causes another incarceration. *History* (in the lexicon of statistics) is the threat that events occurring at the same time as the treatment could cause the observed effect. For example, an observed effect might be due to an event that takes place between the pre and post-test such as may be the case if a subject in a self-efficacy study was offered an attractive position by another employer. *Maturation* is the threat that naturally occurring changes over time could be confused with a treatment effect. People change and events occur naturally over time, and the ways they change may have implications for the dependent variable. For example, the perceptions of people may change as they grow older, experience new things, get hungrier or stronger, etc. Another threat related to time is *attrition* which is the threat that some subjects withdraw from the treatment before it is completed, e.g., if a high pressure treatment is administered, causing only resolute people to remain in the group. Another time sensitive threat is *testing* which is the threat that exposure to a test can affect scores on subsequent exposures, which can be confused with a treatment effect. Some subjects may become sensitized to the aims of a test. For example, familiarity with a test may enhance performance because items are remembered; or (as I personally know) in a weight-loss experiment, weighing someone may cause the person to try to lose weight prior to the next weighing. With all the above known (temporal) threats to statistical validity, quantitative researchers would do well to embrace the epistemology and methodology of qualitative research which has greater capacity to uncover aspects of temporality.

Conclusion

In this paper, I have been speaking as though research can be neatly described as either qualitative or quantitative. However, I do not wish to convey the idea that qualitative research is a well defined, homogeneous area of study – the situation is far from coherent. Although the field is sometimes defined in the negative (as not quantitative), the wide array of qualitative research traditions has been described as a “fuzzy set” (Madill & Gough, 2008, p. 254) of methods clustered in different ways, with “a dazzling array of methodological choices... and a complex

amalgamation of metaphors, paradigms, techniques and procedures that are primarily united by their nonstatistical orientation” (P. Prasad, 2005, p. 3). The same can be said of the quantitative category. There are a huge number of alternative approaches all under the umbrella of positivism. Even within categories, numerous options exist. For example, within the category of multiple regression, one can address research questions with various assumptions around temporality by using either standard multiple regression, sequential regression or stepwise regression. If the researcher makes the judgment that data are nested (for example, students nested in classes which are nested in schools which are nested in district school authorities) then hierarchical linear modelling is a further option with its own set of temporal assumptions.

To adequately deal with my research question, I have run the risk of reifying the quantitative-qualitative divide as a punctuated dichotomy. However, it is worth noting that even in methods traditionally viewed as qualitative, the distinction is not always clear. Numerically-aided phenomenology is one example. I particularly enjoyed the following quote from Fielding & Schreier (2001, p. 43): "even in this type of (postmodern) ethnography, practitioners recognise that all methods impose perspectives on reality by the type of data that they collect, and each tends to reveal something slightly different about the same 'symbolic' reality". This turns postmodernism upon itself. Postmodernism is an ultra-subjectivist research tradition that is characterized by pluralism, fragmentation and a fair amount of whimsical rhetoric. Surely there is room in that tradition (and other traditions) to usefully add quantitative pieces to the mix.

Objectivist conceptions of temporality imply that research subjects all experience time in the same way. This may be seen as a necessary assumption to underpin efforts at generalizability. Subjectivists on the other hand (applying qualitative techniques) understand the passage of time from the point of view of the experience of it. Augustine argued that time is experienced subjectively, existing only “in retrospection, in the moment, and in anticipation... time has no extension other than our immediate experience of it; thus, the measurement of time is only possible as the human mind stores sense perceptions in memory” (Cunliffe et al., 2004, p. 269). This is a similar sentiment to the quote from Ricoeur (1984) used as the opening for this paper. I believe that mixed methods and triangulation help researchers to be more effective and approach their data with healthy scepticism with regard to temporality. Obviously, it is most helpful when social science does not attempt to ‘prove’ hypotheses while ignoring temporal evidence (as in Alice in Wonderland; verdict first, evidence later). The art of research is enhanced by interrelating data from different sources with imagination, and by identifying interesting problems and good study methods. Concern for the validity of research is common in all research traditions. We all want to get it right, even if we have to acknowledge that there may be different versions of what is right. Shadish, Cook & Campbell (2001, p. 478) express the view that qualitative researchers strongly accept the need for validity, and state that the similarity of fundamental research values “are surprisingly more common than some portrayals in the qualitative-quantitative debates suggest and because they demonstrate an underlying unity of interest in producing valid knowledge that we believe is widely shared by most social scientists”.

I will end by referring to what Kritzer (1996, p. 25) calls “the pleasure of the statistical text”. The idea of research as a source of pleasure should be recognized explicitly. Kritzer says that research has a puzzle-like quality: “we do not want it to be too easy (because it will become

boring), but we do want to have a chance of reaching something resembling closure.” The chances of this are far greater if one accepts that all research methods share some fundamental similarities, and contribute valuable temporal insights. Kritzer says that data are fun, sometimes confirming what we know; or as in the earlier mentioned cat’s cradle remark from Donna Haraway, research can also deliver some serious surprises. Consideration of temporality has the potential to deliver creative exploration of the quantitative-qualitative debate. As researchers we struggle to explain the past. Our explanations are “another act of interpretation in a different time and space” (Cunliffe et al., 2004, p. 275). We stand to make a greater contribution to Management and Organizational Studies with temporality shedding light on the interdependence of quantitative and qualitative research.

References

- Abbott, Andrew. (1992). From Causes to Events: Notes on Narrative Positivism. *Sociological Methods & Research*, 20(4), 428-455. doi: 10.1177/0049124192020004002
- Aldrich, Howard. (1988). Paradigm Warriors: Donaldson versus the Critics of Organization Theory. *Organization Studies*, 9(1), 19-25. doi: 10.1177/017084068800900106
- Aminzade, Ronald. (1992). Historical Sociology and Time. *Sociological Methods & Research*, 20(4), 456-480. doi: 10.1177/0049124192020004003
- Bryman, Alan. (1984). The Debate about Quantitative and Qualitative Research: A Question of Method or Epistemology? *British Journal of Sociology*, 35(1), 75-92.
- Burke, Peter. (2012). *A Social History of Knowledge II: From the Encyclopédie to Wikipedia*. Cambridge, UK; Malden, MA: Polity.
- Burrell, Gibson, & Morgan, Gareth. (1979). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*. Burlington, VT: Ashgate Publishing Limited.
- Clark, Peter, & Rowlinson, Michael. (2004). The Treatment of History in Organisation Studies: Towards an 'Historic Turn'? *Business History*, 46(3), 331-352.
- Cunliffe, Ann L., Luhman, John T., & Boje, David M. (2004). Narrative Temporality: Implications for Organizational Research. *Organization Studies*, 25(2), 261-286. doi: 10.1177/0170840604040038
- Durepos, Gabrielle, Helms Mills, Jean, & Mills, Albert J. (2008). Flights of fancy: myth, monopoly and the making of Pan American Airways. *Journal of Management History*, 14(2), 116-127. doi: 10.1108/17511340810860249
- Fielding, N., & Schreier, M. (2001). Introduction: On the Compatibility between Qualitative and Quantitative Research Methods. *Forum: Qualitative Social Research*, 2(1). <http://www.qualitative-research.net/index.php/fqs/article/view/965/2106>

- Foucault, Michel. (1982). *The archaeology of knowledge; and, the discourse on language*. New York: Pantheon.
- Griffin, Larry J. (1992). Temporality, Events, and Explanation in Historical Sociology: An Introduction. *Sociological Methods & Research*, 20(4), 403-427. doi: 10.1177/0049124192020004001
- Haraway, Donna J. (1994). A Game of Cat's Cradle: Science Studies, Feminist Theory, Cultural Studies. *Configurations*, 2(1), 59-71. doi: 10.1353/con.1994.0009
- Haraway, Donna J. (2006). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20th Century. In J. Weiss, J. Nolan, J. Hunsinger & P. Trifonas (Eds.), *The International Handbook of Virtual Learning Environments*. Netherlands: Springer.
- Howe, Kenneth R. (1988). Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard. *Educational Researcher*, 17(8), 10-16. doi: 10.3102/0013189x017008010
- Huselid, Mark A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of management journal*, 38(3), 635-672.
- Kidder, Louise H., & Fine, Michelle. (1987). Qualitative and quantitative methods: When stories converge. *New Directions for Program Evaluation New Directions for Program Evaluation*, 1987(35), 57-75.
- Kritzer, Herbert M. (1996). The Data Puzzle: The Nature of Interpretation in Quantitative Research. *American Journal of Political Science*, 40(1), 1-32.
- Madill, Anna, & Gough, Brendan. (2008). Qualitative research and its place in psychological science. *Psychological Methods*, 13(3), 254-271. doi: 10.1037/a0013220
- Moran-Ellis, Jo, Alexander, Victoria, Cronin, Ann, Dickinson, Mary, Fielding, Jane, Slaney, Judith, & Thomas, Hilary. (2006). Triangulation and integration: processes, claims and implications. *Qualitative Research*, 6(1), 45-59.
- Munslow, Alun. (2010). *The Future of History*. London: Palgrave MacMillan.
- Nash, Roy. (2002). Numbers and Narratives: Further Reflections in the Sociology of Education. *British Journal of Sociology of Education*, 23(3), 397-412.
- Prasad, Anshuman, & Prasad, Pushkala. (2002). The Coming of Age of Interpretive Organizational Research. *Organizational Research Methods*, 5(1), 4-11.
- Prasad, Pushkala. (2005). *Crafting qualitative research: working in the postpositivist traditions*. Armonk, N.Y.: M.E. Sharpe.

- Ricoeur, P. (1984). *Time and narrative* (K. McLaughlin & D. Pellauer, Trans. Vol. I). Chicago, IL: University of Chicago Press.
- Rotermund, HH, Engel, W, Kordesch, M, & Ertl, G. (1990). Imaging of spatio-temporal pattern evolution during carbon monoxide oxidation on platinum.
- Shadish, William R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- TED Talks. (2012). Susan Cain: The power of introverts:
http://www.ted.com/talks/susan_cain_the_power_of_introverts.html?quote=1378.
- Weatherbee, Terrance. (2012). Caution! This historiography makes wide turns: Historic turns and breaks in management and organization studies. *Management & Organizational History*, 7(3), 203-218. doi: 10.1177/1744935912444356
- Weatherbee, Terrance, & Durepos, Gabrielle. (2010). *Dar'wren'ian Evolution in Management: Implications of the Ethics of Epistemology for the Scholar*. Paper presented at the Academy of Management Proceedings.
- Whitener, Ellen M. (2001). Do "high commitment" human resource practices affect employee commitment? A cross-level analysis using hierarchical linear modeling. *Journal of Management*, 27(5), 515-535.
- Zikmund, William G., Babin, Barry J., Carr, Jon C., & Griffin, Mitch. (2010). *Business Research Methods* (8th ed.). Mason, Ohio: South-Western Publishing.

ASB 2013
Antigonish, Nova Scotia

Graham Haynes (student)
Gerald Schwartz School of Business
St. Francis Xavier University

**WE'VE BEEN BUSY "PUTTING OUR HOUSE IN ORDER:"
A TALE OF ONTOLOGICAL POLITICS AT THE NOVA SCOTIA MUSEUM OF
INDUSTRY¹**

The following paper presents an analysis of ontological politics through the development of the Museum of Industry, one of the 27 Nova Scotia Museums in the province of Nova Scotia, Canada, and the struggles between actors to achieve their particular vision of the museum and of history.

Introduction

1986-1998, 12 years: That was the amount time that passed between the first developments of the Nova Scotia Museum of Industry [NSMOI] and its final completion. This timeline spans three major museum hiring processes, two director changes, two closure threats and one major change in ownership - from provincial control to private interests and then, 13 months later, back again to the Province. From this emerged a differentiation in vision regarding (a) what the purpose of the museum of industry should be and (b) how the museum should arrive at that state. These conflict/resolution relationships were navigated through the mobilisation of human and nonhuman actors, that is, through politics, and revolved around the multiple ways in which the museum was enacted in varying sets of practices.

The following paper tells a story of ontological politics (Mol, 1998). It is a story of the development of the Museum of Industry, looking primarily at the development of the museum under the two most influential actors – first, the province of Nova Scotia, and second, the Friends of the Nova Scotia Museum of Industry society [FNSMOI Society]. The former conceived the idea of the museum from a failing industrial base in Nova Scotia. They saw the museum as an outlet for education and heritage and valued the preservation of an older way of doing things. The latter planned to run the museum as a world-class interactive site, seeing museum etiquette and theory as primarily an unimportant barrier, but more importantly, envisioned the museum as a visitor attraction and a sustainable business model. This story details the struggle (politics) between the two to enact their visions of interpretation that would ultimately shape the Province's history.

¹This project would not have been possible without the support of the StFX Summer Student Research Internship Program, a component of the St. Francis Xavier University Strategic Research Plan.

I use a combination of postmodern historiography and actor-network theory (ANT) to engage the past. As such, I will follow the actor's involved in the development of the museum and its change in focus through the years, and more importantly, the politics of those changes. This paper proceeds in three sections: In the first section, I unpack the literature which informed my findings; next, I explain my methodology and the data collection procedure; and finally, I share the findings that came from my examination of ontological politics at the Museum of Industry in Stellarton, Nova Scotia, Canada.

Literature Review

The “Historic Turn” in Management and Organizational History (M&OH)

Akin to the scientist slant of Taylor prevalent in the 1960s, Organizational Studies had largely been dominated by a scientific approach. This has meant that many theoretical developments have essentially ignored the socio-cultural factors that contextualize their research (Üsdiken and Keiser, 2004). However, a push led by Zald in 1993 challenged and changed the field by calling for the reconceptualization to a humanistic as well as socialistic study. Coming from within that framework, scholars in the sphere of Management and Organizational History have, in recent years, claimed there is need for a “historic turn” in organizational theory (Booth and Rowlinson, 2006; Durepos and Mills, 2010). Such considerations are especially relevant to understanding the state and purpose of the NSMOI where the historical precedents and vision/motivations are pertinent to the development process. However, there has been considerable scholarly discussion regarding to what extent history should be added back into organizational theory – to which three approaches have been posited. First, the supplementarist position, which adheres to the view of organization theory as a social scientific study, and merely adds history as another contextual variable, alongside other variables such as national culture (Booth and Rowlinson, 2006). Second is the integrationist approach, best summed up by Zald's call for a focus on the ‘intersection’ and the ‘conjoining’ of historical analysis and the study of particular organizational forms and processes (Zald 1993: 520). An integrationist seeks to enrich organization theory by developing links with the humanities, including history, literary theory and philosophy, without completely abandoning a social scientist orientation (Kieser, 2004: 619). The final approach, which most resembles the one utilized herein, is the reorientationalist approach. This approach is bred from the larger movement to postmodernism by seeing the social science-based approaches as problematic just as the natural sciences are (Kieser, 2004). This is done by problematizing the base of understanding –this is part of a larger and fundamental re-direction in organizational studies away from the social sciences (Üsdiken and Kieser, 2004: 324). In this sense, we consider historical (past) events as context and process, not as a variable. This postmodern approach to history is akin to the work of Jenkins (2003), which I draw on in this paper.

Postmodern Historiography

In the following paper, I address history from a postmodern lens (Prasad, 2005) in line with the writings of Keith Jenkins and other postmodern historiographers. This approach to history assumes the epistemological construction of all knowledge. History then, is a very different thing from “the past” – it is positioned and constructed with ideologies and interests. Thus, Jenkins asserts that every history holds within itself an implicit philosophy of history (Jenkins, 2003, 25). Using this conception of history with the NSMOI in mind, the interpretation of artifacts, the way in which they are presented, even the very choices of what pieces of the collection to exhibit are a piece of analysis. We therein assert that there is no real, objective truth. The consequence of this approach is that it allows for a critical analysis, and the possible reappraisal of the metanarratives involved a “history” of “the past” This sort of analysis can be used to authenticate the final consumer’s experience. As Hewison states, (1987:47) “the question then is not whether or not we should preserve the past, but what kind of past we have chosen to preserve, and what that has done to our present.” Because the context of this study is the museum, I now turn to a review of extant literature on museum studies.

Museum Studies

The way in which the past is depicted as “history” at museums has caused scholars increasingly to see the institutional study as a “fertile theoretical field” (MacDonald, 1996: 6) in that they “as staging grounds (Annison, 1986) for many questions which are also at the hearts of debates in social and cultural studies” (Macdonald, 1996:2-3). Particularly to anthropology and sociology, museums, as Richard Handler (1988) notes, turn culture into an object - giving it’s own reality via objects and their interpretations (Macdonald, 1996). It is in this construction of culture that many scholars have revealed cultural assumptions and political motivations in the analysis of museum collections (Hewison, 1987). However, critical studies of museum exhibition are problematic in that, as Gordon Fyfe notes, “in many cases the analyst’s reading is not acknowledged as a particular interest positioned act of interpretation.” (Macdonald, 1996: 4-5) Indeed, Macdonald notes of the museum studies scholars that “...all seem to share a conviction that much of the older theorizing of museums is too mono-dimensional and insufficiently nuanced to account for the space specificity or the complexity of the museum” (Macdonald, 1996: 4). Herein lay the two-fold reason ANT is well-suited for museum studies. First, it overtly presents the author’s problematization, thereby avoiding the unreflexive attitude Fyfe (1995) refers to, and secondly, it applies a broad based semiotic approach to analysis that grants it the necessary tools to decipher complexity in the sense referred to by Macdonald (1996). This theoretical “fit” with actor-network theory has produced well developed studies on topics such as the linguistic development of heterogeneity) (Hetherington, 1999) as well as social memory using museums) (Urry, 1996). In the case of the Nova Scotia Museum of Industry, it will be a tale of ontological politics (Mol, 1998).

Methodology

Actor-Network Theory

Actor-network theory materialized in the social studies of science and technology literature in the late 1980s and early 1990's and evolved primarily from the work of Bruno Latour (1987, 1988), Michel Callon (1986) and later John Law (1999). As it is widely called today, "ANT" can not be summed up simply and quickly.

The actor-network. The essential subject of analysis in actor-network theory, the actor-network, can be derived quite basically, as a network built from the relations between actors. First, however, what is an actor? As stated by Mol, "An actor does things. It, he, she, acts." (Mol, 2010: 255). An actor is any element which bends space around itself, makes other elements dependent upon itself and translates their will into the language of its own (Callon and Latour, 1981: 289). "Common examples of actors include humans, collectives of humans, texts, graphical representations, and technical artifacts." (Mol, 2010: 257)

The concept of an actor-network can be explained plainly by an analogy: when going about doing your business, for example, driving your car, there are a lot of things that influence how you do it. "For instance, when driving a car, you are influenced by traffic regulations, prior driving experience and the car's manoeuvring abilities... All of these factors (or actors) are related, or connected, to how you act. You do not go about doing your business in a total vacuum but rather under the influence of a wide range of surrounding actors" (Hanseth, 1998). Therefore, the act of driving and all the influencing factors must be considered together, in totality. An act must be linked with all of its influences - this is exactly the utility provided by the term actor-network. The theoretical difference of an actor-network, however, is that the elements in it, the links, can be both of technical and non-technical origin. To use the cited analogy, it is not only the car's capacity to drive, but also the rules of the road and your driving training. There, ANT unearths and analyzes the heterogeneous nature of actor networks. "The crucial analytical move made by actor network writers is this: the suggestion that the social is nothing other than patterned networks of heterogeneous materials." (Law 1992: 381) This is a radical claim because we are saying the "social" also includes any other material, from any different source, (heterogeneous) not simply humans. For example, look at the material world the live in – the text we read as written by an author is reliant upon the printing press, the postal system, even perhaps the computer it was written on. Actor-network writers suggest that these networks shape reality itself. (Law 1992)

In practice, each new study not only stands within the ANT-tradition, but also relates to wider discussions about the topic it explores. One text may be in dialogue with epistemology, the next with aesthetics, etc. (Mol, 2010: 262). In particular, I explore the ontological politics behind

the development of a multimillion dollar museum and its programming using actor-network theory. This, I will argue, involves the translation of actors' interests.

Translation. Actors, as stated earlier, have interests. They seek to align the interests of other actors with their own in an attempt to form a network with a uniform way of seeing and therefore doing (Callon and Latour, 1981). Translation is the process that leads to the formation of an actor-network, it is in this process that an actor-network is created. It should be noted, however, that there is no "creator" of any actor-network, but a network of actors that interact and struggle in the creation of a network (Callon and Latour, 1981). As Callon (1986) stated, this process of interest alignment (translation), has four parts, or moments. "Although these can occur in any order, often overlap, and are rarely clear in the moment" (Callon, 1986: 224).

- 1) **Problematization:** The first moment of translation pertains to a focal actor that defines identities and interests of other actors, making them consistent with its own interests, and establishes themselves as an obligatory passage point [OPP]. In doing so, they render themselves indispensable to the process of translation (Callon, 1986). The OPP is a situation that has to occur in order for all the actors to satisfy the interests that have been attributed to them by the focal actor.
- 2) **Interessement:** The second moment of translation involves a process of convincing other actors to accept the definition that the focal actor imposes and stabilizing their identity to fit the problematization. This is called interessement – the alignment of interests to the focal actor (Callon, 1986).
- 3) **Enrolment:** Enrolment is the moment that another actor accepts their interests as defined by the focal actor. "If consensus is achieved, the margins of manoeuvre of each entity will then be tightly delimited" (Callon, 1986). This is not a permanency, but a momentary state, that can be extended and lasting or only achieved momentarily.
- 4) **Punctuation:** Through deconstruction, ANT helps us to see that the actor-networks we usually perceive to be static and concrete (organizations, institutions, etc), are dynamic and always in flux. However, to keep from endless complexity, we often draw on networks that act as one, therefore concealing all the politics of those actors crucial to their formation. ANT scholars refer to these as punctualized resources. By continuing to draw on them as a whole, the network is performed, reproduced and ramified (Law 1992: 384-5).

Ontological Politics

In the process of translation, the underlying politics of actor-networks construct reality. Therefore, when we look back at the course of interest-alignment and power associations, we can see interests enacted by actors as they struggle to define what it is that becomes real. The term used to describe this process is ontological politics (Mol, 1998; Alcadipani and Hassard, 2010). Using the language of ANT, by studying the power relations between actor-networks, we can see how history was shape, how things could have been different (Mol 1998). Reality is thus manipulated by these political actors and the expressions of their interests (Mol, 2010). “The goal of ANT research is to describe the process through which diverse actor-networks came to produce and sustain ontology [reality]” (Bonner et al, 2009: 87). In the case of the NSMOI, I will describe the politics that underlie the struggle over how to “see” – or what would become the reality, of the NSMOI. This involves two primary actors: first, the provincial government of Nova Scotia and second, the Friends of the Nova Scotia Museum of Industry Society.

Using ANT to “Do” History

Recent Management and Organizational History (M&OH) literature has featured a call for further use of history in research data collection. Notably, some scholars have chosen ANT to answer this call. (Holmstrom and Stalder, 2001; Durepos and Mills, 2008) For example, Bonner et al (2009) use ANT to conduct an investigation of the release of personal motor vehicle records. They do this by following the government’s enactments of the so-called “balance” of information privacy/information freedom and how this has continually pushed in the direction of institutions rather than individuals (Bonner et al, 2009). Furthermore, Bruce and Nylands 2011 use ANT in their examination of the rise of the Human Relations School and the translation of interests over time. In the marriage of ANT and history, we have further seen the use of historical analysis applied to museums due to their geoproximal simplicity as a site of history (Hetherington, 1999). The same method will be featured in the research to follow: examining the history of the development of the Nova Scotia Museum of Industry as a site of ontological politics – where the existence of a multiplicity of versions of the past has serious consequences in the overall direction, completion and function of the museum.

Data Collection

Archival Visits. Three data collection days were logged at the Nova Scotia Museum Library Archive in Halifax, Nova Scotia on May 23rd, June 6th, and June 13th in which over 300 data sets were collected and analyzed, averaging 8 documents each.

Interviews. Interview protocols were constructed and two interviews were held with retired employees of the NSM complex. A snowball sampling method was used by which our archival data collection acted as initial respondents. Our “respondents [were] obtained from information provided by the initial respondent” (Zikmund, 2003: 741). The interviews were transcribed verbatim and entered into the database.

Web Searches. Furthermore, web searches were conducted using the name of the 27 museums that currently make up the Nova Scotia Museum Complex (of which the NSMOI is now a part). These searches queried the names of the 27 museums in their respective local newspapers, national news media sources (CBC & the Globe and Mail) and Google. This was done to document critical events as well as gauge media’s depiction of the museums and to find the perceptions of the sites in Nova Scotia and Canada. Relevant articles were saved in PDF format and stored on a database.

A Tale of Ontological Politics at the Nova Scotia Museum of Industry

Interest Translation in the Development of the Museum of Industry

Central to ANT is the idea that individual actors struggle to enact their own interests via mobilizing the support of others (Law, 1999). In the case of the NSMOI, the development of the museum brought together very different actors and thus entirely different interests and motivations. What was agreed upon by these actor-networks? They wanted a museum. However, what the museum was to look like, and how it was to be performed – that was up to focal actor: the one who took the lead in the struggle, who translated others’ interests and enrolled others in their plan. In the story of the story of the NSMOI, the focal actor would be a group of individuals who would come to be known as the Friends of the Nova Scotia Museum of Industry Society.

This section of the paper is organized as follows: we begin by establishing the background and punctualization of the primary actor-networks involved in the shaping of the Nova Scotia Museum of Industry. This includes the Nova Scotia Museum Complex [NSM], local business associations, the nearby citizen population and finally the Society. Following this, by invoking ANT, particularly Callon’s constructivist sociology of translation (Callon, 1986), I will demonstrate how the Friends Society, in a time of financial hardship in Museum development, asserted themselves as the only possible option to continue operation, making themselves an obligatory passage point [OPP] and thus becoming the focal actor. They arranged to use the decommissioning or “mothballing” funds to open the museum and had plans to make it “world class” (FNSMOI Society, 1993a). In doing this, they told every other actor-network what they wanted to hear, therein translating other’s interests into their language. With this, actor-networks were enrolled in their identities and the Society was granted ownership of the museum.

Next, I will discuss the differential in museum vision and enactments between actors, leading to a fatal rift in the enrolment, a second OPP and ultimately failed translation. Before I delve into the translation, I will first introduce the primary actor-networks that had an effect on the Nova Scotia Museum of Industry, and how their interests became punctuated (singular).

Interests of the Nova Scotia museum complex. The Nova Scotia Museum Complex [NSM], with as many as 25 separate Provincial museums under its wing, (MacIsaac, 1994) was formally presented with the proposal for the establishment of a Museum of Transportation and Technology (many minor name changes were considered in its development) in 1974 by a group of businessmen in Pictou County who “noticed that the machinery and effects of these industrial activities and the railway were disappearing. They approached the Board of Governors of the Nova Scotia Museum to request a museum be built dedicated to Nova Scotia’s industrial and transportation heritage” (The Nova Scotia Museum, 2013). In 1982, following several studies, the planning committee decided the need to begin physically compiling the artifacts already obtained, now valued at over 3 million dollars (Frame, 1982a). They also agreed to commission an industrial researcher to continue to compile the collection (Frame, 1982b) which led to the hiring of the museum’s first curator in 1986 and the sod-turning of donated lands in May of 1988 (The Nova Scotia Museum, 2013).

The planning committee did this with the mandate to justify and store items that are otherwise “in danger of destruction or irreversible damage if not taken into custody” (Frame, 1982b). This focus on the collection continued throughout the early 1990’s when basic building construction was completed and is visible in the mandate of the Museum of Industry, which was to “to collect, preserve, research and interpret the industrial history of Nova Scotia” (Stevenson, 1993a). This can also be seen in their decision to hire primarily professional staff for preservation, whereas work such as exhibit development was contracted out (Stevenson, 1993a). Nonetheless, as the newly elected Liberal government enacted province wide spending cuts, museum development was severely impeded and the NSM was forced to consider “mothballing” the museum. This means using the remaining operating money to close the museum until more operating and capital funding could be found (MacDonald, 1993). As one can expect, there was a serious backlash from the general public and pressure was put on the NSM and its government counterparts to avoid closure. Evidence of the backlash is seen in the wealth of news articles published in late 1993 with evocative titles such as “One thoughtless chop” (Ryan, 1993). Some articles went so far as saying “It [the museum] is in danger of never opening and becoming a singularly ironic monument to failure” (McDougal, 1993) and even single out political actors: “How can Education Minister John MacEachern, who represents the working people of Glace Bay, so indifferently discard the heritage of his own riding?” (Ryan, 1993). The political parties involved, including the Nova Scotia Museum Complex itself, had interests in the museum being opened. They wanted it to be a facility for preservation in memory of an older way, but, primarily, they wanted the museum to stay on track to open as scheduled, (Ryan, 1993), not to close. They were looking for answers.

Interests of nearby businesses and trade associations. Tourism was and continues to be a primary industry and economic driver in Nova Scotia. In the name of increases in tourist numbers and spending, local businesses had interests also intertwined in the opening of the Museum of Industry. Included in the media firestorm in late 1993 regarding the museum's closure was the Pictou County Tourists Association president talking about the impact of closure, saying "to shut down the museum project so close to completion is ludicrous" (Taylor, 1993). Other associations who either sent letters to the premier or otherwise spoke out include the Pictou County Development Commission, Pictou County Chamber of Commerce, Stellarton Downtown Business Association, Pictou Business Improvement District Commission, New Glasgow Waterfront Development company and others (Stevenson et al, 1993). These interests for an open museum were fired by an economic impact study commissioned in October, 1993, saying the museum would see 100,000 visitors in its first year, with a future potential of 150,000 – of which 70% would come from outside Pictou County and 30% from outside Nova Scotia (Doane Raymond Company, 1993). Thus, businesses were interested in the necessity of accommodation, food service, goods purchase and other spending from the tens of thousands of possible visitors. They wanted to see a museum open - one that would attract as many people as possible, at that. In an economy of decline, they too, were looking for answers.

The interests of the nearby citizenry: donors, employees, families and others. With the government having invested over 17 million dollars at the time of closure talks, (Taylor, 1993) the looming threat of job loss from museum-preparation and lost opportunity from visitor spending, the general public simply wanted to see the museum opened. This can be seen in the abundance of letters to members of parliament, news articles, radio interviews and more regarding the possible closure of the museum (Unknown, 1993) and the degree of anger in their prose. There were also rumours of plans to "disperse irreplaceable artifacts and sell the building", as stated in a news article, that obviously gave cause to a wave of fear and anger amongst the donor community (Ryan, 1993). The greater community of Nova Scotia residents were looking for answers or, better yet, a solution to the funding dilemma in the opening of the NSMOI. They wanted a museum open with their friends', families' and neighbours' heirlooms, but most of all, they wanted an open museum - that is exactly what the Friends Society would offer.

The interests of the focal actor: the friends of the Nova Scotia museum of industry society. In late 1993, when the Nova Scotia Museum had to consider mothballing the NSMOI due to government cuts, a society known as the Friends of the Nova Scotia Museum of Industry came forward as a concerned collective of businessmen and political candidates, with a shared interest to save the museum – a black-boxed actor-network (FSNMOI Society, 1993b). However, this complete conceptualization of the Society hides all the actor struggles that brought it to this punctualized point. On the official website for the NSMOI, it states that in 1974, a Pictou County group of businessmen presented the BOG with a "proposal to study the establishment of a museum of transportation and technology in Nova Scotia" (The Nova Scotia Museum, 2013). However, in the history written in 1993 by the director of the NSM at the time, Candace Stevenson, wrote to the minister saying "the concept...evolved as a result of discussions between the Board of Governors of the Nova Scotia Museum and Mr. William Sobey...in 1974, Mr. Sobey

offered the Province 5.5 acres of land as the site of an industrial museum” (Stevenson, 1993a). Why does this matter, one might ask? Mr. Sobey’s son, Donald, who finalized the land transaction with the Nova Scotia Museum, was the President of the Society (FNSMOI Society). Indeed, in 1993, when the Friend’s offer came forward, it was initially referred to internally as the “Donald Sobey Plan” (Stevenson, 1993b). The real conceptualization of the Museum can be found in William Sobey’s speech at the sod-turning (he was such an important actor in development he delivered the sod-turning speech) in 1988: “It was ten years ago that I first approached Lynton Martin in his office when he was director of the NSM. I suggested that I would like to see a mining museum built in Stellarton and that I was prepared to provide an excellent site of suitable size for that purpose. Mr. Martin replied, I’m afraid, Mr. Sobey, I cannot become very excited over a Miner’s Museum for Stellarton as we have a Miner’s Museum in Glace Bay. However, if you wished to excite me, why don’t we discuss a Museum of Transportation and Industrial Technology?” (FNSMOI Society, 1993b).

After William Sobey’s death in 1989, his sons Donald and Frank (especially Donald) continued their father’s vision in a generally passive fashion throughout museum development. This was of course only until 1993, when provincial budget cuts left the possibility of closure (FNSMOI Society, 1993c). On September 9, 1993, Donald Sobey and a group of town mayors and leading Nova Scotia business owners agreed that “the group would pursue the development of a plan to maintain the operations of the Nova Scotia Museum”, looking to use “...a society [as] the vehicle for the development of the plan” (FNSMOI Society, 1993d). Thus the Society was created, with the Sobey family interests, to open a museum. They subsequently created a management plan, business plan and market strategy to sustain the museum (FNSMOI Society, 1993b). When they went public with their heroic message in 1993, headlines beamed with titles such as “Business leaders aim to save museum”, (Graham, 1993), calling the Society an “interested group of corporate and municipal officials” (White, 1993). By deconstructing the Society interests, we get a different picture than the way they were presented at the time.

Furthermore, the society’s vision was different from that of all the other actors – they saw the museum as being a “world class institution”, (MacLellan, 1994) but more importantly a business, with revenues and expenses and profits. They were the first to conceptualize the museum with a souvenir shop and food outlets (FNSMOI Society, 1994b), calling their museum a “highly commercialized operation which is highly marketable and capable of substantial revenue generation” (FNSMOI Society, 1993e). Through the very presentation of a business, marketing and management plan, one can see the interests involved in the Sobey (Society) vision. Most notably, one can notice the interest differential in a memo written from the Society in their search for a loan near the end of their tenure (as we will find) in which they state that “Funding the five year road to self-sufficiency is difficult for both the Society and the Government, however, a sustainable museum will pay dividends for many years to come” (Murray, 1995). Their vision had to do with money, not artifacts or people. Nonetheless, the public message forwarded by the society was exactly the message the relevant actor-networks wanted to hear – they would open a world class museum using minimal funds. And in this, they would set themselves as an obligatory passage point (Callon, 1986).

Moments of Translation

The society problematizes the museum plan. In a time when the museum was slated to close, the Society presented their intent in a management plan with cost alternatives to the “options” the NSM has – either mothball the museum or levy an annually decreasing operating subsidy (less than the cost of decommissioning) to the Friends, who will be given control to operate the museum (FNSMOI Society, 1993a). With this, it would seem the only way to keep the museum open, that is, to have the outcome that all actor-networks are looking for, was to follow *their* plan, thus making themselves indispensable to the situation. This situated them, as Callon (1986) would remark, as an obligatory passage point.

Interessement: the society tells everyone what they want to hear - in their language. Aligning the interests of the relevant actor-networks is known as interessement in Michel Callon’s (1986) sociology of translation. It is, at this point, worth reviewing the interests of the involved actor-networks in the museum. First, the NSM Complex wanted an open museum and one that preserves the past; the business community was looking for a museum of grandeur to bring in maximum tourists; lastly, the public desired an open museum for economic gain, social gain, political gain and the like. To them, the Friends proposal of a “world class” museum, opened using only mothball money, seemed the perfect solution (FNSMOI Society, 1993a). The Society publicly “intend[ed] to work with the Nova Scotia Government, municipalities, corporations, and the general public for the efficient and cost effective management and operation of this extensive facility” (White, 1993). The actor-networks were ready to help because after all, this represented the alignment of their interests – this was what they wanted, and according to the problematization as set by the focal actor (the Society), this was the only way forward.

Enrolment and mobilization: resources needed - after all, this is what you wanted: As the relevant actor-networks accepted their roles, the Society was able to control and manage the museum as they saw fit because of the actor-networks they had persuaded to accept the roles attributed to them. When the interest-aligned actor-networks accept their identities that the focal actor attributes them, these actor’s are “enrolled” in their plan – this is the third step in Callon’s (1986) translation process. The Society had now successfully enrolled the networks: if they need money, the NSM gave it to them, if they needed volunteers, the public was willing to help, if they needed support, businesses were willing to pitch in their help. After all, these were the roles they accepted because this is what they wanted and they were told this was the way to achieve it. *The focal actor has them enrolled.*

Thus the Society’s conceptualization of the museum, with the language of world-class commercialization and business knowledge was adopted and used by the other actor-networks in different settings (LeBlanc, 1995). As these roles were accepted and enacted, the identities are

“mobilized” in different settings, the fourth and final step of Callon’s (1896) sociology of translation.

A second problematization: rising from the conflict of interest. This enrolled development continued for months until a problem occurred. As mentioned above, there was discontinuity between how the NSM and the FNSMOIS envisioned the actualization of the museum: to the NSM, it was to be a place of heritage; to the Society, a sustainable business model. The problem was not in that the Society did not see the Nova Scotia Museum’s interests (realistic heritage preservation) as important or vice-versa, but rather that the two parties had conflicting primary interests. In early 1996, as the Society is now struggling to run the museum on their anaemic budget (MacEachern, 1996), two problems with the Friend’s enactment of the NSMOI were brought to the NSM: one, that the catalogued records for thousands of artifacts do not exist, (Jardine, 1996) and second, that the museum had cut the last of it’s professional staff (Stevenson, 1996a) despite the fact that artifacts continued to be worn, damaged and repaired (Baillie, 1996). For the first matter, this went to show the importance differential discussed earlier. “I felt compelled to write this memo as there have been times when I have been dismayed at the lack of attention the collection has received...” writes a previous Registrar, “the direct management of the collection was placed on the back burner [under the Friend’s management]” (Jardine, 1996). Yet as we know, the *Province’s* number one agenda, indeed the reason for the conceptualization of this museum, was the collection (The Nova Scotia Museum, 2013). As for the second matter, that of the collection’s preservation, the executive director of the museum, Candace Stevenson, takes a hard line on the issue, and in doing so illustrates the NSM interest in the collection. She then writes a letter to the Society’s director saying,

“We must therefore consider the collection to be static at this time i.e. no additions to it, no restoration work, no loans, no use of items except as they are now used/displayed – unless or until I have been consulted” (Stevenson, 1996a) and in a letter to the Province regarding future funding, she says “they will receive no funding unless they hire someone with museum experience as a director or curator. Furthermore, they cannot hire anyone without someone from the NSM on the hiring boards” (Stevenson, 1996b).

In this, the museum problematizes the Society’s actions and asserts *itself* as an obligatory passage point!

Failed enrolment: not simply a financial partner - loss of trust, loss of museum: As the Society struggled to maintain the enrolment of the NSM, they were also dealing with the inability to gather funding to continue to run and develop the museum (MacEachern, 1996). They were unable to receive the significant private funding they wanted because, according to them, they need a solid footing, a financial promise from the Province, before investors were willing to bite (Sobey, 1996). At this point in 1996, according to the Friends, “they are in year 2 of a 5 year plan to make a World Class Museum. They will require additional monies from the Government in order to get to the end of the 5 years and have a World Class Museum” (Jolly, 1995). The latter

quote surfaces the language by which the Friends aligned the interests of the relevant actor-networks.

However, now frustrated and trying to translate the Society's interests, the NSM responded by disallowing the loan without more documentation, the Executive Director going as far as saying "I am increasingly concerned that the Board [of the Society] has no idea of where they want to go with this institution or what a museum really is" (Stevenson, 1996b). In a following memorandum sent from the NSM, a fourth option is presented to the Province which reads: "choose to run the Museum again as a directly-managed museum in the Nova Scotia Museum System" (MacEachern, 1996). The initial problematization of the Society had finally been disrupted, the enrolment now a failure with the most important actor-network involved. In June of 1996, the NSM was granted permission to resume direct management of the museum, thus ending the direct involvement and soon, the existence, of the Friends of the Nova Scotia Museum of Industry Society (Province of Nova Scotia, 1996). One year following this, in retrospective, the NSM Director now in charge of the NSMOI Candace Stevenson writes in a letter to a local MP stated that "as you know, we've spent the better part of this fiscal year 'putting our house in order' at the Museum of Industry" (Stevenson, 1997).

Conclusion: The Struggle to Sustain Ontology

A Case of Ontological Politics

Tracing the history of the NSMOI illustrates a case of politics and, more specifically, a case of ontological politics. This case demonstrates that history is always subject to and an outcome of politics. Actors were engaged in struggles over how to see the museum and also the history that it is dedicated to commemorating. This shows the flexibility of history, wherein actors struggle to enact their particular vision the museum and the history that it is dedicated to commemorating – the museum could have looked different, a different history could have been told. In the Friend's control of the museum, the Society dictated the development of exhibits and, concurrently, the interpretation of them as well. As stated by a retired NSM senior staff member,

"We used to get a car and drive up there to Stellarton and stay there for a night or two, sit down with John Hault (past NSMOI Director) and sort of talk about the shape of the exhibits. It was a planning kind of thing... we were trying to plan exhibits, that was our mandate... *We tried to have some input into the shaping of the museum; I don't think we were entirely successful*" [italics added] (Interviewee, 2013).

Through the museum's punctuation, the politics of actors (as they struggled over how to see the past) are concealed and the current (extant version of) interpretation stands in as the dominant authoritative history, as if there is one uniform version of the past. As stated on the museum

website, the Nova Scotia Museum of Industry tells *the story* [italics added] of work and workers” (The Nova Scotia Museum, 2013). Yet as we can see from our analysis, “the” history is really just “a” history. There are other histories that could have been told and a different museum that could have been established. Ontological politics is the process by which actors shape reality – history is always the outcome of politics.

References:

- Annis, S. (1986). The museum as a staging grounds for symbolic action, *Museum*, 151, 168-171.
- Baillie, C. (1996, February 19). [Letter to Aubrey DeWolfe]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Bonner, B., Chiasson, M., Gopal, A, (2009). Restoring balance: How history tilts the scales against privacy. An Actor-Network Theory investigation. *Information and Organization*, 19, (2), 84-102.
- Booth, C. and Rowlinson, M. Management and Organizational History: Prospects. *Management & Organizational History* 1(1), 5-30.
- Bruce, K. and Nyland, C. (2011). Elton Mayo and the Deification of Human Relations. *Organizational Studies*, 32(3), 383-405
- Calás, M.B., Smircich, L. (1999). Past Postmodernism? Reflections and Tentative Directions. *Academy of Management Review*, 24(4), 649-672.
- Callon, M. (1986). Some Elements of a sociology of translation: domestication of the scallops and the fisherman of St Brieuc Bay. J. Law (Ed.), *Power, Action and Belief: A new sociology of knowledge?* (Vol. Sociological Review Monograph 32, pp. 196-233). London: Routledge Kegan Paul.
- Callon, Michel, and Bruno Latour. (1981). Unscrewing the big Leviathan: how actors macro-structure reality and how sociologists help them to do so. *Advances in social theory and methodology: Toward an integration of micro-and macro-sociologies*, 277-303.
- Doane Raymond Company (October, 1993). [Economic Impact Study for the Nova Scotia Museum of Industry]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Durepos, G., Mills, A.J., Mills, J.H., (2008). Tales in the manufacture of knowledge: Writing a company history of Pan American World Airways. *Management & Organizational History*, 3(1), 63-80.
- Durepos, Gabrielle, and Albert J. Mills. *Anti-history: Theorizing the Past, History, and Historiography in Management and Organization Studies*. Charlotte, NC: Information Age Pub., 2012.
- FNSMOI Society (1993a, October 29). [Society Management Plan Proposal for the NSMOI]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- FNSMOI Society (1993b, September 22). [Business & Marketing Plan for the Museum of Industry]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.

- FNSMOI Society (1993c, February 23). [Special 8 page publishing inside the Advocate Newspaper, Pictou, NS]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- FNSMOI Society (1993d, September 9). [Minutes from the Society Gathering pre-registration]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- FNSMOI Society (1993e, September 22). [Nova Scotia Museum of Industry Programme Outline]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- FNSMOI Society (1994, May 12). [Friends of the Nova Scotia Museum of Industry Society Memorandum of Association]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- FNSMOI Society, (1996, June 21). [Memorandum of Agreement for MOI between Society and Province]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Frame, R. (1982a, February 18). [Minutes of the Museum of Transportation and Industry Planning Committee]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Frame, R. (1982b, January 13). [Minutes of the Museum of Transportation and Industry Planning Committee]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Fyfe, G. (1995). The Chantrey Episode: Art Classifications, Museums and the State, c1870-1920. S. Pearce (ed.) *Art in Museums*, London: Athlone, 5-41.
- Graham, M. (1993, September 23). [News article from the Daily News, Truro, NS]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Handler, R. (1988). *Nationalism and the Politics of Culture in Quebec*. Madison: University of Wisconsin Press.
- Hanseth, O. (1998). Socio-technical webs and actor network theory. Retrieved on July 10, 2013 from <http://heim.ifi.uio.no/~oleha/Publications/bok.6.html>
- Hetherington, Kevin. From Blindness to Blindness: Museums, Heterogeneity and the Subject. *Actor Network Theory and after*. By John Law and John Hassard. Oxford. England: Blackwell/The Sociological Review, 1999. 51-73. .
- Hewison, Robert. *The Heritage Industry: Britain in a Climate of Decline*. London: Methuen London, 1987.
- Holmstrom, J. and Stalder, F. (2001). Drifting technologies and multi-purpose networks: the case of the Swedish cashcard. *Information and Organization*, 11(3), 187-206.
- Interview with previous Nova Scotia Museum employee, (2013, May 24).

- Jardine, T. (1996, January 26). [Memo to Candace Stevenson]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Jenkins, Keith. (2003). *Re-thinking History*. London: Routledge. .
- Jolly, S. (1995, September 19). [Letter to Premier John Savage and Hon. John MacEachern]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Kieser, A. (1994). "Why Organization Theory Needs Historical Analyses – And How This Should Be Performed." *Organization Science* 5(4), 608-620.
- Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard UP. .
- Latour, B. (1988). *The Pasteurization of France*. Cambridge, MA: Harvard University Press.
- Law, J. (1992). Notes on the Theory of the Actor-Network: Ordering, Strategy and Heterogeneity. *Systems Practice*, 5, 379-93.
- LeBlanc, K. (1995, September 6). [Letter to Museum Staff from the Pictou County Tourist Association]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- MacDonald, D. (1993, September 20). [Memo sent to Fay Lee]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- MacDonald, S. (1996). Introduction. *The Sociological Review*, 1(1), 1-18.
- MacEachern, J. (1996, May 26). [Memorandum to the Nova Scotia Government Cabinet]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- MacIsaac, Catherine, (1994, September 7). [News article entitled from the Hants Journal, Windsor, NS]. Nova Scotia Museum Library Archive, Halifax, Nova Scotia, Canada.
- MacLellan, D. (1994, June 18). [Newspaper article from the Evening News, New Glasgow, NS]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- McDougal, T. (1993, December 30). [News article from the Daily News, Truro, NS]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Mol, A. (1998). Ontological Politics. A Word and Some Questions. *Sociological Review*, 46(S), 74-89.
- Mol, A. (2010). Actor-Network Theory: sensitive terms and enduring tensions. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 50(1), 253-269.

- Murray, S. (1995, September 19). [Letter to Sandy Jolly]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Province of Nova Scotia, (1996, July 2). [Agreement between the FNSMOI Society and the Province for the MOI]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Ryan, J.H. (1993, September 26). [News article entitled from the Daily News, Truro, NS]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Sobey, D. (1996, May 2). [Letter to Bernie Boudreau]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Stevenson, B.J., Clarke, D., Hardy, D., Hetherington, L., Fleming, A. (1993, November 8). [Packet of letters written to various ministers on various dates]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Stevenson, C. (1993a, June 23). [Memo sent to Armand Pinard]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Stevenson, C. (1993b, September 8). [Letter to John MacEachern]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Stevenson, C. (1996a, February 1). [Letter to Aubrey DeWolfe]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Stevenson, C. (1996b, March 13). [Letter to Hon. John MacEachern]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Stevenson, C. (1997, April 7). [Letter to Hon. Robbie Harrison]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Taylor, W. (1993, September 13). [News article from the Pictou County Advocate, Pictou, NS]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- The Nova Scotia Museum. (u.d.). History. *Museum of Industry*. Retrieved July 15, 2013, from <http://museum.gov.ns.ca/moi/en/home/about/history.aspx>
- Unknown Author (1993a, October 19). [Transcription of Radio Interview regarding the Museum of Industry]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.
- Urry, J. (1996). How societies remember the past. *The Sociological Review*, 1(1), 44-65.
- Üsdiken, B. and Kieser, A. (2004). Introduction: History in Organizational Studies, *Business History*, Vol 46, No 3, 321-330.

White, R. (1993, September 22). [Newspaper article from the Pictou Advocate, Pictou, NS]. Nova Scotia Museum Library Archive, Halifax, NS, Canada.

Zald, M. (1993). Organization studies as a scientific and humanistic enterprise: Towards a reconceptualization of the foundations of the field. *Organizational Science*, Vol 4, 513-528.

Zikmund, W. G. (2003). *Business Research Methods*. Thompson South Western: Ohio.

FIFTY SHADES OF PLAY: MONOPOLY™ AND THE SPIRIT OF CAPITALISM

While the recent trend towards the ‘gamification’ of business and organization is forging a new intersectionality between play and games with work and organization, play and games as a subject of research within the MOS literature is limited. While games, as a sub-domain of play, have been an area of study it is normatively approached from a narrow and instrumental perspective. This paper takes an historical look at the intersectionality of games and work from a historical and socio-cultural perspective. ANTi-History is first used to re-assemble a history of the board game Monopoly™. A mediological analysis is then used to investigate the material, ideational, and ideological elements of board games in general and to surface the relationship of Monopoly™ and its direct antecedents with work and organization in particular.

Introduction

While *play* has long been historically recognized as a key and distinguishing feature unique to human civilization, play did not become a focus of serious scholarly interest until the middle of the twentieth-century (Malaby, 2007). At that time Huizinga argued that play, along with cognition and tool-use, lay at the very core of modernity and was therefore a central characteristic of the cultural development of the West (Huizinga, 1949).¹ However, despite this penetrating observation, in the intervening years since play has been predominantly viewed simply as a leisure activity. An activity or sphere of life (Moss Kanter, 1997) somehow separate from daily life (Caillois, 2001; Huizinga, 1949) and standing apart from work and organization in (Dandridge, 1986; Rhodes & Parker, 2008). As a consequence the general study of play has rarely been granted the legitimacy of serious independent scholarly or intellectual endeavor (Pisac, 2013). Rather, when it has been considered at all within mainstream social science, play has been a relatively minor and secondary subject of study. Similarly, *games* as a form of play have also largely avoided any comprehensive scientific investigation.²

Most cultural-anthropological research on play and games conceptualizes these activities in one of two ways; first as a material practice defined in terms of the productivity of its activities, and second, play as an activity which is purely symbolic in nature. In the former approach play is positioned as non-productive or a *wasteful* use of one’s time (Caillois, 2001). Latterly, play as a symbolic, ceremonial, or ritual cultural activity is viewed as being largely

¹ Huizinga analytically distinguishes both Homo Sapiens (Man the Thinker) and Homo Faber (Man the Maker) from Homo Ludens (Man the Player).

² The terms play and game remain ambiguously and multiply defined across the social science literature. In its modern instantiation play seems to be defined more so by what it is not (e.g. it is not work, it is non-productive) rather than what it is. Similarly, the term game tends to be viewed as a structured (e.g., rule or domain bounded) form of play.

representationalist in nature. Representations which are then relegated to a secondary status of other more primary or culturally meaningful activities (Geertz, 1972).³ It has been these two dichotomous approaches which have dominated the theoretical and methodological study of play for the last half-century (Malaby, 2009). However, this circumstance is now changing and the relatively recent explosive growth of gaming as a leisure activity within late twentieth-century popular culture has acted as a catalyst for renewed practitioner and scholarly interest.⁴ With this recent attention there has been calls for counter-debate to the classic view that everyday life and play are separate and distinct spheres (Beck & Wade, 2004; Kane, 2004).

As a result there has been increasing attention paid to the intersectionality of games with organizations and work.⁵ The fundamental assumptions which have pre-figured the cultural roles and functions of play and the separation of work and play into non-commensurable spheres are being rethought (Malaby, 2007; McGonigal, 2011). Play and games are now being actively discussed within the popular business press as businesses ‘gamify’ themselves in order to attract and engage with an electronically connected and globalized society (see, for example Belsky, 2012; Chatfield, 2010; Edery & Mollick, 2010; McGonigal, 2011; Sullivan, 2011). But, despite the recent re-emergence and renewal of play and games as pan-life-world constructs, the study of games still remains largely absent in the management and organizational studies (MOS) literature. A review of MOS work over the last fifty years demonstrates little variance with the normative pattern of relative absence of interest in the study of play. While research involving games has a greater presence than that of play, this research may be seen to approach its subject from one of three orientations; instrumentally, pedagogically, or linguistically.

The instrumental or utilitarian literature is historically founded in the efficiency imperatives of scientific management (Taylor, 1911) and the administrative necessity for complex organization wide decision making. A combined approach which arose out of the operational research initiatives taking place in the United States during the Second World War (Boulden, 1970; Swan, 1950-1952). This is the oldest and most developed stream in MOS (see for example Cangelosi, 1965) and the research is derived from these economic, mathematical/probabilistic, and game theoretic foundations (Mirowski; Scribeber, 1958; Swan, 1950-1952; von Neuman & Morgenstern, 1944). The primary aim of this research has been to explore the utilitarian potential of games, models, and simulations in their employment as predictive decision aids in production and manufacturing (Goetz, 1959) and for organizational decision making; primarily executive or managerial (Lewin & Weber, 1969).

Closely related to the instrumental treatment of games has been the investigation of their application for pedagogical purposes (see for example Leemkuil & de Jong, 2012; Neuhauser,

³ These are the two primary, but it may be argued both dichotomous and reductionist, perspectives from which play and games are conceptualized. In the case of the former, play and games were assumed as either non-work or non-productive activities (see Caillois, 2001); and in the case of the latter, as representative of cultural processes of meaning-making (see for example Geertz, 1972).

⁴ For example, the advent of the concept of ‘gamification’ or the use of game or game-like elements with productivity applications on the World Wide Web or in mobile applications for business (Zichermann & Cunningham, 2011). While games and game theory have long been used in business and management education, i.e. business games or simulations for decision making or for experiential training and education, there has been an increased focus on their potential due to the extensive penetration of computers in the educational systems and the evolution of the world wide web and distance or web-enabled learning (Kapp, 2012).

⁵ Two signals of the ‘sea change’ in the study of play and games have been the formation and publication of several new journals including *Game Studies* (2001) and *Games and Culture* (2006).

1976). This literature is grounded in cognitive and behavioural learning perspectives, i.e., the use of play and games for experiential training and education for students or practitioners of management (Wolfe, 1973).⁶ While the concept of play and games as a sociocultural and performative activity is most closely approached by this literature (see for example Kark, 2011) the aims of much of this research are still largely aligned with those found in the instrumental stream. That is, games employed as a utilitarian tool for managerial or organizational purposes.

Finally, in the linguistic frame to play and games, the terms and concepts are employed somewhat differently than those of the culturally defined activities referred to by Huizinga. Here, these terms are conceptually employed to study both the nature of language and its use at work (i.e., language-games). The language-games approach to play comes at these practices from a postmodern/post-structural basis (c.f. Lyotard, 1984; Wittgenstein, 1965). Relatedly, in terms of language use, the terms play and game(s) are also commonly used metaphors. While the former tends to focus on language as political rhetoric, e.g., the jargon surrounding business and markets (Rindova, Becerra, & Contardo, 2004), the latter is employed to draw social parallels from the domain of play and games so as to highlight certain aspects of phenomena in business and organizations. The use of play and game as metaphors is quite prevalent in both the academic and practitioner literatures. Examples include; organizational governance and leadership *games* (Westphal, 1998), the *game* of internal organizational politics (Donnellon, Gray, & Bougon, 1986), or to describe the normative *rules-of-the-game* for career advancement in occupations of professions (Prasad, 2013; Raelin, 2008). Even when viewed in the aggregate all of these approaches still leave MOS with a general dearth of effort in this area (for exceptions see Dandridge, 1986; Meier Sorenson & Spoelstra, 2012).

While this brief review highlights the diversity of research in the utilitarian, pedagogic, or linguistic study of play and games, most of this research still does not investigate the socio-cultural intersectionality between play and games with work and organization. Even less study is conducted from a socio-historical perspective.⁷ This is considered problematic as a socio-historic perspective teaches us that the taken-for-granted nature of managerial and scholarly beliefs, e.g., the separation of leisure from organization and work, are likely to be historically contingent.

The Socio-Historical Gaze

Most social science scholars still rarely attend to the metaphysical assumptions which underpin the theory and methods they employ (Van Maanen, Sorensen, & Mitchell, 2007). Recognition of this was largely brought to light by Burrell and Morgan's (1979) historical tracing of sociological theory. While the surfacing of these assumptions spawned much debate, coined the 'paradigm wars' (Jackson & Carter, 1993; Willmott, 1993), discussion of metaphysical assumptions again subsided in mainstream MOS. While history was implicit in the paradigm

⁶ See, for example, journals such as *Decision*, *Decision Management*, *Decision Sciences*, and similar publications grouped into the subject areas of Operations or Management Sciences.

⁷ Game Theory is only very loosely related to the broader cultural concept of play and games. While Game Theory has its roots in games of chance, it is a reductionist view of games taken from the narrow perspectives associated with mathematics/probability studies and the economics of decision utility and risk. Consequently, in this study, while the common root of origin for play-games and Game Theory's is acknowledged it will not be treated with further

discussions it was an unreflective use of history that did not theoretically or methodologically tie itself to the broader socio-historic context within which much of the positioning of the debate was conducted. This was/is a circumstance symptomatic of the largely ahistorical approach of much of the work in our discipline. In this regard MOS has also failed to surface and debate the more fundamental problematic of the historical dimensions of our own knowledge (Jacques, 2006; Parker, 2004). Of such concern is the ahistorical nature of MOS research that several scholars have strongly called for the inclusion of more history in organizational studies (Clark & Rowlinson, 2004; Kieser, 1994). There have since been numerous additional arguments justifying greater theoretical and methodological engagement between MOS, history, and the humanities (Booth & Rowlinson, 2006; Zald, 1993, 1996). While this is now beginning to take place (Weatherbee, Durepos, Mills, & Helms Mills, 2012) the turn to history remains nascent with much of MOS remaining untouched by any *socio-historic* gaze; including the arena of play and games.

Monopoly™, ANTi-History and Mediology

One particular game, the well-known game of Monopoly™, has been chosen as an ‘ideal’ game-object of study. This has been done for pragmatic, methodological, and theoretical reasons. Pragmatically, Monopoly™ is currently one of the world’s most enduring board games, i.e., it has a significant historicity, and is well known, i.e., it has a significant historical presence. Methodologically, therefore, there is a great deal of historical facts or trace which may be re-assembled into a history. Theoretically, as Monopoly™ is a cultural artifact which has successfully transitioned multiple changes in the conceptualization of play, work and organization within the context of modern Western culture, it affords the opportunity of using an historical, versus ahistorical, approach to theory development.

This study traces the evolution of Monopoly™ by employing two complementary historical methodologies; ANTi-History (Durepos & Mills, 2012) and Mediology (Debray, 1996, 2004). As ANTi-History is a relatively recent approach in MOS (Durepos & Mills, 2012; Durepos, Mills, & Weatherbee, 2012; Mills, Weatherbee, & Durepos, 2013), and since, as far as this author has been able to determine, mediological approaches have yet to be used within a MOS context (The first comprehensive presentation of the mediological method is still very recent, see Debray, 2004), the following sections will briefly summarize why these methods are appropriate for an historical study of this type.

An ANT framework was chosen as the ontological assumptions of Actor-Network Theory (Latour, 2005), the heterogeneous nature of actor-networks comprised of both human and non-human elements, very strongly resonates with the concept of material play or game; i.e., games as a performative social object that is comprised of both ideational and material elements. Drawing upon mediological techniques, specifically the concepts associated with the processes of *transmission*, *organized materiality* and *materialized organization* (Debray, 1996, 2004), allows Monopoly™ to be analyzed in ideational (symbolic) and performative (material cultural practices) terms. The use of ANT and mediology in this way permits the surfacing of both the

materiality and the performativity of play-as-board-game while. Materiality is defined here as the process of design, construction and organization of game materials (e.g., board, playing pieces, etc.) by various actor-networks. Performativity is both the playing of the game (e.g., enacting the rules and conduct) as well as those activities that are not play but are related to the physical and ideational presence of the game within an historical context (e.g., applying for patents, marketing and selling the board game, etc.). The methods of ANTi-History are used for the re-assembly (Mills et al., 2013) of a history of the gaming activities of various actor-networks. A re-assembly focused upon the intersections between historical situated actor-networks, board games, and work and organization over time to place them into context. Mediology highlights those elements needed to explain how materialization and ideation are linked in cultural practices from an historic perspective.

Debray's work (1996, 2004), for which he coined the neologism *mediology*, is similar to Actor-Network Theory (Latour, 2005) in so far as it is not viewed as either a science or a theory (Debray, 2004). Instead, like ANT, mediology is conceptualized as a method or approach to study the social. In counter-poise to ANT, however, a mediological project is broader and seeks to understand the socio-material conditions which permit the transmission of culture and how society both produces and reproduces itself (Vandenberghe, 2007). Thus the central empirical questions for mediological approaches are derived from Marx and Althusser; i.e., how do ideas become a material force and under what material conditions do some ideas survive while others do not? Put differently, how do ideas, which are ephemeral constructs expressed within a culture both linguistically or symbolically, come to have the power to enroll, persuade, or otherwise capture the attention of individuals and collectives such that they enact them and make them real? For mediologists, the answer lies in the reconceptualization of the modern (and modernist!) understanding of communication and its relation to our notions of ideation, materiality, and performativity.

A mediological perspective starts by splitting communication into two related but analytically and empirically distinct processes. These processes are the communication of information and the transmission of information. When analytically separated communication may be seen as the transport or movement of information in *spatial* terms. Thus, information is communicated from one place to another place. Transmission, on the other hand, may be seen as the processes of transport or the movement of information in *temporal* terms. Thus, transmission systems and processes act to connect or relate information temporally, from one time to another time. It is only in very modern communications where the message and medium appear synchronous in space and simultaneous in time. It has been the technical capability of immediacy in our communications, which have become commonplace and taken for granted, which has masked the collapse of the spatial and temporal dimensions of communication into one. Modern society has forgotten that instantaneous global communications is a historically recent phenomenon, e.g., only with the advent of electrical/electronic technologies and that communications have not always been so. Throughout most of the history of civilization communication and transmission were distinct and separate processes. To use a now 'quaint' example, consider the letter (information communication) and the postal service (information transmission). Mediology's conceptual separation of communication from transmission permits a more refined and socio-historical understanding of how ideas move through space and time and

how societies first form and then reproduce themselves. In cultural terms, it is the communication of ideas which allows for social formation; i.e., collectives to cohere or networks to form, while it is the transmission of ideas which permits the collectives or networks to endure and have continuity over time (Debray, 2004).

A mediological perspective also attends to the performative nature of the transmission of information, knowledge, or ideas. While cultural studies looks largely to semiotics, symbolism and the media (i.e., communication), it rarely attends to the materiality of technology as technology (i.e., transmission) as it maintains its focus on the social over the material. While technological studies try to understand the use and influences that technology has in social and cultural terms, its focus is centered on the material at the expense of the social (Vandenberghe, 2007) and the role of technology in the formation of collectives is absent. It is in the overlapping Venn-space which conjoins communication and transmission, ideation and materiality, that mediology sees social, cultural and technological hybrids in performative terms. It is communication and transmission in performance/praxis/practice which “makes the spatial production of society through communication of information possible and that, by storing memory, contributes to its temporal reproduction through the transmission of culture from generation to generation.” (Vandenberghe, 2007, p. 26). To use Debray’s terminology, it is the actions of *materialized organization* (the network, the institution) which are generative of *organized materiality* (the material objects) which translate (or transforms) ideation into material force. Put otherwise, performativity is the materialization of ideas through organization.

It is in this moment, when the material organization of an idea is manifest and understood as a material force that mediology moves beyond ANT as the materials must be embedded in a network of relations to sustain them as “[s]tability does not inhere in materials themselves.” (Law, 2009). Mediology marries the ideational or symbolic ‘word’ with the heterogeneous and material nature of the ‘world’ in seeking to understand the dynamics of ‘the production of signs and the production of events’ (Debray, 1996, p. 17). However this is not to say that mediology is dismissive of ANT. For it is the performativity of actor-networks which may now be seen as working to organize the symbolic into the material with the results of their praxis the organized materiality which acts as a vector for the reproduction and transmission of knowledge. In this sense, mediology may be considered as a materially-discursive orientation that views ideas as manifest in material form. A form which imparts material force, and hence durability, and allows objects and systems of objects to transmit ideas over time. From this one can see that ideas need materialization and praxis (organization) in order to be durable. It is this notion which separates the communications and transmissions functions of classical communications theory (Debray, 1996) and allows for ontology and epistemology to be collapsed into the material force. What Barad (2003) conceptualizes as the intra-action or performative transformation of the word into the world; where both the word and the world, the symbolic and the material, are placed into a mutually constitutive relation. When methodologically used in combination ANTi-History and mediology permits the identification and analysis of actor-network performativity, organized materiality and materialized organization in a socio-historical context. Further, it also highlights both the ideational and material elements and how they are related in actor-network performativity.

In order to surface how actor-networks work to produce and disseminate organized materiality in such a way as to perpetuate or reproduce the symbolic or ideational elements from one historical context to another we must first historically situate and then re-assemble a socio-history of board games. The following sections will, in sequence; contextualize the social nature of board games, re-assemble a history of Monopoly™, and describe the intersectionality of games and work in the socio-historic context of America from which Monopoly™ emerged.

Board Games in History

Our knowledge of play is as old as human historical and socio-cultural reckoning and there is much anthropological and archaeological evidence for play materialized as board games. Game boards, game pieces, and inscribed rules of play appear to have existed in parallel with the development of civilizations and for most of our collective history board games were locally produced artefacts created by individuals. These artefacts spread in ways similar to that of oral culture moving from person to person, community to community, and sub-culture to sub-culture. How board games were made in pre-industrialized civilizations would vary based upon the local availability of materiality, creativity or design choice, and social purpose. The rules associated with the games would also vary from community to community, and culture to culture as board games were first and foremost a variable social folk practice (Flanagan, 2009).

In Western culture, from ancient Greece through Rome and into Northern Europe, there is a great deal of material evidence for board games (Kurke, 1999; Whittaker, 2004, 2006). The historicity of board games is to be found in other non-Western cultures as well. The antecedents of the popular Chinese game of *Go* can be dated even before 500 BC (Eberhard, 1968), while the modern form of chess, as known within the Western world, had its origins in the distant past of India. From there it spread to Russia, Africa, and China before eventually reaching Europe around 1000 AD (Murray, 1913). Similar style board games were played as early as 5000 BC in Mesopotamia, Egypt, and Persia (Flanagan, 2009). Historically then, board games appear as ubiquitous and historical as our cultures themselves. The historical evidence would also seem to invoke a *prima facie* case that the material forms of ancient board games, including playing pieces with boards or maps, were the cultural precursors to the modern board games so common in the nineteenth and twentieth centuries. While this may be true on a material basis – it is not the case for the social and cultural purpose of games. Such a conclusion would be in error. For like play (Huizinga, 1949), the place, role, and purpose of board games must be historically examined and contextualized in socio-cultural terms to be understood. Fortunately, board games themselves are amenable to lending a hand in this respect.

As materialized “cultural texts” (Sugarman, 1994, p. 323) board games can give us insight into their social functions as board games often reflect the dominant interest and concerns of specific historical and socio-political contexts (Hofer, 2003; Kurke, 1999). They can be interrogated to separate out the universal themes of human play from those themes that are particular to different cultures at different times. In this way, individual board games yield insight into a culture’s particular perspective at one point in time, and when viewed collectively may yield insight into the stability or change of those interests over time. So our first entry into the investigation of board games and their relationship with the then contemporary

conceptualization of work and organization is to be through an historical lens. A lens focused most especially on the purpose, place, and role - the 'who' and 'why' - of board games in general and of Monopoly™ in particular.

Re-assembling Monopoly™

There has been a great deal of effort expended in the historical study of Monopoly™. It is a complex story that has unfolded a little at a time over a period of almost 75 years. This is not unusual given its historicity, its popular status, and its global ubiquity in the Western world of leisure and games. However, of pragmatic necessity, due to the focus of this study as well as the limitations of space, a fully detailed history of all the various actors and events involved will not be presented. Readers are instead referred to the several very detailed histories on Monopoly™ (see Anspach, 1998; Axelrod, 2002; Hofer, 2003; Kennedy, 2004; Orbanes, 2006). In order to situate the various actor-networks within their historical context(s) a re-assembly of the historical trace is described in the sections which follow. This re-assembled history is a summarized and synthesized description of the facts (persons, events, locations, and dates) sourced from the publications cited above. Where other sources beyond these have been used they are referenced separately. The relevant actor-networks are briefly described in rough chronological order for the sake of clarity.

The Georgists and the Ideologue: Henry George was a Progressive economist who developed an alternative economic theory in response to the growing monopolism of capitalism in the United States in the 1800's (Collier, 1979; George, 1912). George's theory demonstrated the relationship between land, labour and capital as they related to a governments' ability to generate tax revenue for the purpose of investing in the economy and social development. George's economic theory was premised upon the concept that the private ownership of land or the holding of monopolies on natural resources generated wealth but did not add to the productive capacity of the economy nor contribute to the economic needs of the community (Heilbroner, 1999). He saw the rise of monopolies in the United States as a disturbing trend as the value of monopolistic holdings only increased as a function of scarcity meaning that much wealth accumulation was essentially economically non-productive as the value of holdings increased as a function of scarcity or speculation. George saw this as having two important and adverse effects on the economic well-being of the nation and its people. First, because the increase in value of monopolistic holdings was non-productive it did not contribute to the growth of the economy overall. Second, the failure of the government to tax land ownership and the monopolists of natural resources shifted the burden of social development (e.g., the costs of community commons improvements) onto those who could least afford it; the renters and non-monopolists.

George saw this circumstance as producing a vicious circle where increased monopolistic wealth accumulation allowed for speculation and wealth generation while productive work/workers bore the cost of the tax burden as it moved away from the increasingly wealthy *rentiers* to those of lower economic strata. This in its turn served to increase the total non-productive wealth being accumulated by monopolists and worsening the condition of the productive segments of the economy. George firmly believed that this only served to further

exacerbate the gap between the rentiers and the renters, between the rich and the poor. A conclusion drawn from what George had observed in the pattern of industrialization and urbanization of the United States throughout the latter part of the 1800's. To resolve this circumstance George developed and advocated a Single Tax policy; a land tax that would reduce non-productive wealth accumulation and which would more equitably benefit everyone in society through equitable taxation policy (George, 1912). His work was considered a very progressive theory in its day - his text went on to sell more than 3 million copies after its publication in 1879 (England, 2010) – and the importance of his theory continues to be discussed today over a century later (Hoff, 1991; Shoup, 2004). In its day, his ideas resonated very strongly with many in America and abroad. One of these persons was Lizzie Magie.

Lizzie Magie, the person who designed and constructed one of the immediate material predecessor games to Monopoly™, was a social progressive and an ardent proselytizer of the economic theory of Henry George. She had been exposed to Georgian social economic theory through her father and her residency in the Single Tax community of Arden Delaware (see below). In order to educate members of society in George's economic theory and to demonstrate the adverse consequences of poor economic policies of government, e.g., allowing for private monopolies, Lizzie Magie designed and developed a board game which she called the Landowner's Game. She applied for and received two patents for the game, the first in 1904 for the original version, and a subsequent patent in 1924. The second was for a variant of the game re-designed to more fully highlight the economic dangers of monopolies (Magie, 1904; Magie Phillips, 1924). These games were explicitly designed to show how the accumulation of wealth through land ownership and private monopolies would lead to an increasing gap between economic *rentiers* and non-landowners. The 1924 variant of Landlords saw Lizzie Magie rewrite the rules to include a second rule set, titled Prosperity, which explicitly showed how the financial and economic differences in outcomes between monopolistic versus anti-monopolistic were based on the player's economic choices. In other words, the Landlords Game had been materially (game board, game pieces) and ideationally (rules of play and the economic model of the game) designed to highlight the then contemporary problematics of the capitalist system and the potential for a Georgist economic policy as a fairer alternative. For Lizzie Magie the game was designed to be "educational in its nature" in order to "illustrate to them [the players] how under the present or prevailing system of land tenure, the landlord has an advantage over other enterprises and also how the single tax would discourage land speculation." (Magie Phillips, 1924). Through personal advocating, word of mouth, and the establishment of the Economic Game Company, Lizzie Magie's efforts saw the spread of the Landlord's Game throughout the Northeastern United States. Much of the spread of the game was also enabled through the cultural practices of 'folk gaming'; a process which was inextricably linked to Lizzie Magie's game.

The Folk Gaming Practice: Prior to the industrialization of America, in the United States as elsewhere, folk games were a dominant form of entertainment. Games would be created and played within localized communities. In the early 1900's the process of folk gaming impacted the evolution of the Landlord's Game in both material and ideational terms as various actors and networks produced and played their own 'folk' versions of the game. These folk gamers were comprised of individuals and groups who had been exposed to the game by Lizzie Magie, the

Ardenites (see below), and other gamers brought together by social acquaintance. For example, members of the community of Arden, an experimental community established upon the Georgist Single Tax philosophy in 1900 (Group, 2013), regularly re-created, i.e., made by hand with materials at hand, and played the Landlord's Game. These versions went either by the names Landlord's, Auction-Monopoly, or Monopoly depending on which rules of play, monopolistic versus anti-monopolistic, were preferred. In this fashion, localized versions of Landlord's were spread from player to player, place to place, and community to community for over two decades with each rendition of the game similar to the one before, yet also varying in material and rules. One of the members of the Arden community who re-created and played the Landlord's game in this way was Scott Nearing.

The Academy: The Progressive, anti-monopolist, and political radical Scott Nearing was introduced to the Landlord's Game while a resident of the Arden community. Whether he was introduced to the game by Lizzie Magie or others in the community is unknown. However, when Nearing left Arden to pursue his education at the University of Pennsylvania, he took the game with him. After completing his PhD in economics he assumed a position teaching at the University of Pennsylvania Wharton School of Finance and Economics from 1906 through 1915 (Collection, 2013; School, 2013). While there he used the Landlord's Game as a teaching tool to demonstrate the anti-social nature and harmful effects of monopolistic economic practices. However, due to his radical social and economic views he was ultimately dismissed from Wharton (School, 2013). After he was fired from Wharton he took a position at the University of Toledo (1916-1917) (Collection, 2013). It is not known if he taught using the Landlord's Game at this institution as his tenure at Toledo was very brief. He was let go from the university due to his openly expressed anti-war sentiments. After his position at the University of Toledo, he became a free-lance journalist and lecturer, publishing articles and texts, and advocating for various social improvements; many of which directly involved educating others of the dangers of monopolistic practices and the social ills caused by the capitalist system of economics. While at Wharton, several other persons, students of Nearing and other fellow students, were exposed to and became interested in the game. One of these was Rexford Tugwell; a student of Nearing's who would himself become an economist. He would go on to teach economics at Columbia University from 1920 through 1936 (University, 2013). Tugwell would also introduce the game to other students at Columbia University as he also used it as teaching tool in class. One of the students exposed to the Landlord's Game was Daniel Layman. Layman would go on to produce and market a variant of the Landlord's Game which was to be called Finance. Finance, as a game, had been stripped of all of its anti-capitalistic 'political' or 'social' sentiments and was designed to be played on purely capitalistic and economic terms. The elements of social critique inherent in the previous versions were 'written' out of the game.

The Monopolists: Charles Darrow, who was proclaimed the 'inventor of record' was the individual who first produced and sold the first version of a board game officially, entitled Monopoly. Darrow had been exposed to the game through the folk gaming community and in classic American entrepreneurial fashion began producing and selling the game in the local markets of the northeastern United States. He had essentially transformed what had been a folk game and folk practice into a production and manufacturing practice. The success of the game's sales drew the attention of Parker Brothers one of the largest manufacturers of board games in

America. Parker Brothers, and later on Hasbro who purchased Parker Brothers in 1991, purchased the intellectual property rights from Charles Darrow and paid him royalties on each game subsequently sold. Eventually, Parker Brothers purchased the rights for all of the variants of *Landlord's / Monopoly / Finance* from Lizzie Magie, Daniel Layman and others. They patented and trademarked the name Monopoly in order to protect their investment. Between them these companies produced, marketed, and sold the Monopoly™ board game worldwide. Today, it is estimated that some 270 million copies have been sold around the globe with over a billion people having played the game (Hasbro, 2013).

Ideology and the Shifting Hegemonies of Play

In socio-historical terms the modern Western notion of childhood and society's role and responsibility towards children is very recent. The contemporary and modernist social conceptualization of childhood is but the latest in a series of developments and shifts arising out of social change stemming from the enlightenment. Childhood and children as socio-culturally defined categories have changed and evolved several times since (Aries, 1962). One line of historical study places the emergence of the social differentiation between adult and child, the root conceptualization of modern-day childhood, as late as the 1600's (Postman, 1985). Adolescence as a liminal category between childhood and adulthood would not appear until the late 1800's and would not fully emerge and stabilize until the early 1900's (Lesko, 1996). Prior to this children were socially considered "as miniature adults whose sole purpose was to become big adults as quickly as possible." (Eiss, 1994, p. 1). This is a significantly different positioning of children in social terms than held in modern Western culture today. The contemporary concepts that we associate with childhood, e.g., play, protection, learning and education, were far different as recent as a century ago. The social meanings and functions associated with play and games as they related to children in the nineteenth century were also very different.

In Western culture, and particularly in America, the 'child' of the 18th and 19th centuries, most especially the male children, were seen as a potential and "a profitable source of labor" (Eiss, 1994, p. 10). Children were future adults and workers; first for the agrarian family and then, with the advent of the industrial revolution, for the factory. At this time childhood, gender, and work were normatively bundled conceptualizations. Males were often the only "economic providers for their families" and so "biological sex" was synonymous with "economic function" and the masculine gender, including that of children, was defined almost exclusively in economic terms (Blair, 1994, p. 308). Male children were to be managed and trained as future labor assets who would, in their turn, become economically productive adults and contributors to society. Only then could they fulfill their economic role as breadwinners for family and the productive backbone of society and the nation (Lesko, 1996). So strongly held were these social norms that "[t]hose males who lack[ed] the mental quality of Industry lack[ed] the external sign by which it [was] known, labor". Children who failed to assume their role as industrious adults in society were considered "economic castrati." (Blair, 1994, p. 309).

So board games from their earliest inception in Western thought, through to America in the late 1800's, had two combined social purposes, play and education (Kurke, 1999). However, the first board games in America emerged prior to the reconceptualization of childhood in the

twentieth century and so most board games were designed for the training of future adults. That is to prepare children for work and organization. As the United States transitioned from a primarily agrarian to primarily industrial society "the American home, [was] no longer the heart of economic production" and as the spheres of work and home life separated, the home became "the center of education, entertainment, and moral enlightenment." (Hofer, 2003, p. 13). Board games were therefore "concentrated on specific, narrow, practical matters" (Eiss, 1994, p. 2) in order to prepare children for their future economic roles as adults. In this context board games taught children how to "take turns, how to follow rules and how to win and lose in the vocabulary of their times." (Sugarman, 1994, p. 324).

While games such as chess and checkers had been imported to America from Europe (and the other antecedent geographies prior to Europe) during the establishment of the country, it was not until the late 1800's that American manufacturers were producing their own original designed-in-America board games (Hofer, 2003). At this time most of the board games produced in the United States reflected American society's overriding interest and preoccupation with religion, entrepreneurship, economics, and industry. Many of the American produced board games reflected America's competitive spirit, technological achievements and the educational and moral development of children (Hofer, 2003). For example, in the Game of The District Messenger Boy (1886) success was defined as avoiding the sins of sloth and dishonesty in order to progress and to become a respected businessman and member of good standing in society. Similarly, in the game The Errand Boy (1891) actions and rule play were designed to highlight the importance of industriousness and moral behaviour which were considered key to career and economic success. These board games reflected the specifics of how children were viewed by society at large (Sugarman, 1994) and were "based on the idea that children should only read [or do] what would improve them." (Eiss, 1994, p. 3). Other board games provided education on the arts (Gems of Art: A New and Instructive Game), geography (The Dissected map of the United States), History (The Historioscope Panorama), language (Anagrams and Other Letter Games), morals (The Checkered Game of Life (1860) and social roles (The Elite Social and Sentimental Conversation Cards), and on religion (Pilgrim's Progress, The Mansion of Happiness, and Grandma's Sunday Game of Bible Questions).

By the late 1800's one of the defining themes in American board games was the accumulation of wealth through economic capitalism, or 'rags to riches' stories. One advertisement for the board game The Monopolist (1885) read, "On this board the great struggle between Capital and Labor can be fought out to the satisfaction of all parties, and, if the players are successful, they can break the Monopolist and become Monopolists themselves." (as cited in Hofer, 2003, p. 84).⁸ So most of the board games, as objects to be performed as edutainment or education in the home, exposed and taught the children of America the primary economic and social values, beliefs, norms and behaviours, and expectations for success in adulthood. As children's games contain the ideological elements of the dominant social order (Limon, 1983), in the case of America, they were educated in morality and trained in industriousness, the economics of capitalism, careerism, and the accumulation of wealth. An ideology passed from one

⁸ Titles, dates and descriptions are sourced from (Hofer, 2003)

generation to another and that would be carried forward and would only grow stronger throughout the twentieth-century.

Ideology: An American Perspective

In the context of the United States' experience, ideology is commonly understood to be a political system of beliefs. Beliefs that are normatively antithetical to the nature of democratic liberal market capitalism in America; e.g. other forms of economic or state control such as communism, fascism, and socialism (Lodge, 1977). As a consequence in the United States ideology is largely viewed in a negative or pejorative sense; likely due to the prominence of ideology read as a central concept of the political 'left' stemming first from Marx's materialism (Marx & Engels, 1947) then Gramscian hegemony (Gramsci, 1971) and latterly Althusser's ideological apparatus (Althusser, 1971). Ideology has been defined as both false belief (Marx & Engels, 1947) and true belief (Mannheim, 1985). Beliefs founded upon, in the case of the former a materialist, and in the case of the latter, an idealist perspective. While ideology has been a core theoretical and analytic concept in political science (Goren, Federico, & Kittilson, 2009; Marx & Engels, 1947) and the sociology of knowledge (Mannheim, 1985), outside of Bendix's work (1956) it has been employed only sparingly within MOS (for exceptions see Alvesson, 1984; Barley & Kundra, 1992; Bendix, 1956; Beyer, 1981; Beyer, Dunbar, & Meyer, 1988; Burnham, 1941; Starbuck, 1982).⁹

In MOS ideology has been variously defined as "logically integrated clusters of beliefs, values, rituals, and symbols" (Starbuck, 1982, p. 3) or "broadly defined ways of understanding that are based on shared values and beliefs." (Beyer & Hannah, 2002, p. 483), or following Bendix as "a system of ideas concerning management that explain and justify managerial authority and that are related to a distinct set of managerial techniques and practices." (Parush, 2008, p. 49). Barley and Kunda define ideology as "a stream of discourse that promulgates, however, unwittingly, a set of assumptions about the nature of the objects with which it deals." (Barley & Kundra, 1992, p. 363). The conceptual use of ideology in this way in MOS has meant that ideology has been bereft of its political interpretations.¹⁰ What is left is a managerialism where ideology has been extended to, and defined in terms of, a system and pattern of thinking whose assumptive and taken-for-granted nature precludes questioning or imagining of other ways of seeing the world (Alvesson, 1984). From this perspective "The 'spirit of capitalism' is therefore an ideology; one that "serves to sustain the capitalist process in its *historical dynamism* while being in phase with the *historically specific and variable* forms that it takes." (Chiapello & Fairclough, 2002, pp. 187, emphasis added). So if ideology, conceptually and ideationally, has remained relatively stable in MOS - even while varying within specific socio-historical contexts - what mechanisms or processes facilitate or contribute to ideological endurance in this way?

One potential answer is to be found by analytically separating out the ideational from the material, or reproductive, components of an ideological system. In other words, the ideation associated with the system from the performativity of the system. Althusser has provided a

⁹ As we see today, at that time board games were a form of "edu-tainment" (Hofer, 2003, p. 20).

¹⁰ More recently, Parush (2008) argues that in management and organizational studies that the management 'fashion' stream of research has essentially replaced what used to be understood as management 'ideology'.

partial answer (1971). In brief, Althusser proposed that the reproduction of any social system was dependent on the dominant ideology successfully being passed from one generation to the next. For Althusser this is achieved through the actions of Ideological State Apparatus (ISA) and Repressive State Apparatus (RSA). ISAs are the institutional mechanisms, such as the system of education which serves to teach the ideational components of ideology, while RSAs, are the mechanisms that maintain the dominant order through action. That is, the performativity of the government and the judiciary when enacting laws and legislation through enforcement by state police and militaries. While ideology tends to be concealed within the actions of the ISAs, it is most apparent in the actions of the RSAs (Althusser, 1971) and taken together this matrix serves to implicitly and explicitly reproduce both the ideological and the structural basis of the social system. In this way ruling hegemonies are perpetuated in both ideational and material forms.

From this perspective it can be seen that board games in general and Monopoly™ in particular have served to educate and reproduce a capitalist ideology within the United States and the capitalist system. However, board games are neither ISAs nor RSAs - so what are they? Moving beyond the Althusserian ISA/RSA matrix allows an argument to be made that board games should be considered a form of Ideological Materialized Apparatus (IMA). IMAs defined as social objects performative of both the ideational and material components of an ideological system which serve to transmit ideology over time. Or, in socio-cultural-material terms, Landlord's/Monopoly™ culturally *transmitted* the idea of capitalism in the form of an *organized materiality* through (Debray, 2004) when played. However, play was dependent, in the first instance, on making. In order for play to happen, the game needed to be both organized and performed. First, the game board, the player's pieces and other paraphernalia, and the rules had to be materially organized. Only once organized materiality was successful could a network be formed for play. The formation of an actor-network itself required two further conditions. First, actors needed to pass through an obligatory point of passage; acceptance of the game world. Second, once within the game world, the rules needed to be performed or enacted. Only then could play happen; otherwise, when not performed as game, the ideological elements of capitalism lay dormant within the game box.

From a very young age we learn to observe and frame the complexities of the world through the use of metaphor. Metaphor thus pervades our language and structures our understanding and interactions with the world from our earliest years (Lakoff & Johnson, 2003). "*The essence of metaphor is understanding and **experiencing** one kind of thing in terms of another.*" (italics in original, bold emphasis is added, Lakoff & Johnson, 2003, p. 5). Perhaps the most dominant metaphor for business in American culture is the 'game' (Clancy, 1989; Donaldson & Lorsch, 1983). The metaphor of business as a game has a long history and many notable businesspersons in the 19th and twentieth centuries often described business as a game or used gaming metaphors to conceptualize business and business practices (Clancy, 1989). Many of them describe their practices as based upon what they learned during their play of Monopoly™ (Axelrod, 2002). The system of metaphors that structure our thinking also serve to highlight and hide elements of the underlying metaphorical concepts we employ in our language and thinking (Lakoff & Johnson, 2003). As the patterns of social behaviour in games informs us that individuals use the bounded game space, the rules of the game, to inform their decision making and actions, game players will often grant the rules of the game higher priority in their

decision making than from any moral reasoning (Reall, Bailey, & Stoll, 1998). By thinking of 'game' as an innocent and leisure activity, as we tend to do today, we have fail to perceive the intersectionality it has and had with work and organization.

The re-assembly of Monopoly™ using ANTi-History and mediology reveals that Monopoly™ was first an ideational based *token*, a quasi-object or *mutable mobile* (Law & Mol, 2001), which was passed within and between folk gaming networks. It was then translated into an *immutable mobile* which had been black-boxed to conceal the underlying ideational/ideological nature of the game (Latour, 2005) imbued into its materiality by Lizzie Magie and others. This effect had come about in the transition of the Landlord's/Monopoly variants as folk games to the mass-produced consumer game Monopoly™. The loosely related networks of the folk gaming community meant that both the ideational and material components of the game spread slowly and were open to variation; depending on the capability and interests of the creator/user. However, when manufactured, marketed, and sold by Darrow and then Parker Brothers, the variable Landlord's/Monopoly folk practice had been stabilized. It had been fixed as organized materiality through a strongly materialized organization; the effect of which was to bind the ideational and material components of the game into an ideological potentiality waiting to be performed over and over again.

While the rules of play could (and undoubtedly were) variable in their application, e.g. house-rules, the materiality of the game board was not as variable – it had essentially become materially immutable. The relative immutability of Monopoly™ was reinforced as Parker Brothers protected Monopoly™ by excluding other similar game ideations and materialities through the use of trademarks, and patents. Combined with the actions of the judiciary which enforced these protections on others (for a very spectacular case of this see Anspach, 1998) the effect was strengthened. Through processes of Debrayan *organized materiality* and the establishment of a strong and coercive network, materialized organization, Parker Brothers created and sustained a monopoly on Monopoly™. In this punctualization Monopoly™ had become both an actor-network and an actant (Latour & Woolgar, 1986). A materialized vector which served to transmit (Debray, 2004) and reproduce a capitalist ideology. An ideology whose fundamentals had been black-boxed (Latour, 2005) and hidden, enrolling others through their 'benign' acts of play. Once the game had become culturally embedded as a stabilized and black-boxed actant, only then could it be materially re-organized into the variants of Monopoly™ we see today (for example, see the many games based upon Monopoly™ at http://www.hasbro.com/monopoly/en_CA/).

The concept of organized materiality and the transmission of ideas, knowledge, and information over time arguably extends Gramscian (Gramsci, 1971) and Althusserian (Althusser, 1971) notions of ideology. A mediological stance can be seen to extend this work and the concept of ideology in three manners. First, mediologists argue that ideology cannot be understood in purely hegemonic ideational terms. In other words, ideology qua ideology has no material force (Marx & Engels, 2004). And without material force, ideology cannot be hegemonic (Gramsci, 1971). Second, and here the extension, without material force ideology cannot be produced/reproduced, even by Althusserian Repressive State (RSA) and Ideological State (ISA) Apparatus. For in order for ideology to be hegemonic and for it to reproduce itself

ideology must first be materialized, organized, communicated and transmitted. Thus, the dissemination and adoption of an ideology is neither solely idealist (Gramsci, 1968) nor materialist (Marx & Engels, 1947) in nature. It is to be found in the processes of *transmission* wherein materialized ideology is transported through time. To return to the Althusserian question, it was the stabilization, punctualization, and transmission of Monopoly™ as a culturally embedded object which produced an Ideological Material Apparatus. An empirical case which demonstrates, to borrow from McLuhan, that sometimes the medium can be the message.

Conclusion

Despite renewed interest and practice of games and work, a closer inspection reveals that rather than being a new phenomenon, play and games have intersected with and been directly related to work and organization for a substantial amount of time. In the North American context, play, games, and work have intersected more often than not. The assumption of the separation of play and games from work and organization is a modernist assumption; one particularly dominant throughout the twentieth century (Malaby, 2007). By following the historical traces of play manifest in the materiality of games we can begin to re-assemble the relationship of play and games with their culturally assigned meanings (e.g., their relations with work and organization) over time. This is viewed as important as games function as social technologies through the ordering logic and relationships set up by the rules of play and games have "assist[ed] in the development of conceptual processes, to invoking ritual, to forging a connection with time and the future, players throughout history have struggled to gain agency and understand uncertainty through game play." (Flanagan, 2009, p. 73).

An historical approach which acknowledges both the materiality of play-as-games as well as their symbolic/ritual cultural role(s) has several advantages. First, it avoids the theoretical problematics of reductionism that have plagued the two classical approaches mentioned previously (Malaby, 2007). Second, the reconceptualization of play as having both material and symbolic/ritual dimensions allows us to investigate the linkages between the cultural or ideational contexts of games and their materiality within and across specific socio-historical contexts. Methodologically this permits a surfacing of the intersections and relations between play and games with work and organization. An historical re-assembly allows the opening of these 'black-boxes' of intersectionality.

Under the rhetorical rubric of *gamification*, games and game elements are expected to increasingly intersect with spheres of activities not classically associated with leisure pursuits. This is the case now with modern communications technology (Zichermann & Cunningham, 2011), the increase of work practices based on games/game elements in organizations and business (Reeves & Read, 2009) and in all ranks of education (Jenkins et al., 2009; Kapp, 2012). Given the rapidity with which *serious gaming* (Ritterfield, Cody, & Vorderer, 2009) and processes of *gamifying* have made inroads into the practices of management and work, the lack of study into the socio-cultural nature of games is viewed as a growing gap; a gap that urgently needs to be addressed by MOS scholars. Hopefully, this study goes partway to delineating and explicating one of the "ideological mechanisms through which the worker becomes accustomed

to accepting the roles, rules, and linguistic usage belonging to the workplace” (Burell and Morgan, 1979 Sociological Paradigms, p. 322-324, as cited in Alvesson, 1984, p. 66).

Finally, strangely, and paradoxically enough, what had been designed as an tool to inform persons about the dangers of particular forms of capitalism and economic exchange through materialization, translation, and stabilization had become a game which served an opposite purpose; one that accentuated all that was originally envisioned as wrong with contemporary capitalism.

References

- Althusser, L. (1971). Ideology and Ideological State Apparatuses (Notes Towards an Investigation) (B. Brewster, Trans.) *Lenin and Philosophy and Other Essays*. London: New Left Books.
- Alvesson, M. (1984). Questioning Rationality and Ideology: On Critical Organization Theory. *International Studies of Management & Organization*, XIV(1), 61-79.
- Anspach, R. (1998). *The Billion Dollar Monopoly Swindle*. Palo Alto: American Printing.
- Aries, P. (1962). *Centuries of Childhood: A Social History of Family Life*. New York: Vintage Books.
- Axelrod, A. (2002). *Everything I Know About Business I Learned From Monopoly*. Philadelphia, PA: Running Press.
- Barad, K. (2003). Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. *Signs: Journal of Women in Culture and Society*, 28(3), 801-831.
- Barley, S., & Kundra, G. (1992). Design and Devotion: Surges of Rational and Normative Ideologies of Control in Managerial Discourse. *Administrative Science Quarterly*, 37(3), 363-399.
- Beck, J., & Wade, M. (2004). *Got game: How the Gamer Generation is Reshaping Business Forever*. Boston: Harvard Business School Press.
- Belsky, G. (2012). Six Reasons Why "Gamification" Will Rule the Business World. *Time: Business and Money*. Retrieved from Time: Business and Money website:
- Bendix, R. (1956). *Work and Authority in Industry*. New York: Harper & Row.
- Beyer, J. (1981). Ideologies, values, and decision-making in organizations. In P. Nystrom & W. Starbuck (Eds.), *Handbook of Organizational Design* (pp. 166-202). New York: Oxford University Press.
- Beyer, J., Dunbar, R., & Meyer, A. (1988). The Concept of Ideology in Organizational Analysis. *Academy of Management Review*, 13(3), 483-489.
- Beyer, J., & Hannah, D. (2002). Building on the past: Enacting established personal identities in a new work setting. *Organization Science*, 13(6), 636-652.
- Blair, S. (1994). "What are you going to make out of this boy?": The Role of Poetry in the "Management" of Boys, 1875-1900. In H. Eiss (Ed.), *Images of the Child* (pp. 307-322): Bowling Green University State Popular Press.
- Booth, C., & Rowlinson, M. (2006). Management and Organizational History: Prospects. *Management and Organizational History*, 1(1), 5-30.
- Boulden, J. (1970). *Interactive Decision Simulation: A Revolution in Management*. Paper presented at the Academy of Management Proceedings.

- Burnham, J. (1941). *The Managerial Revolution: What is Happening in the World*. New York: The John Day Company, Inc.
- Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organizational Analysis : Elements of the Sociology of Corporate Life*. London: Heinemann.
- Caillois, R. (2001). *Man, Play, and Games*. Urbana, IL: University of Illinois Press.
- Cangelosi, V. (1965). The Carnegie Tech Management Game: A Learning Experience. *Academy of Management Journal*, 8(2), 133-138.
- Chatfield, T. (2010). *Fun Inc: Why Gaming Will Dominate the Twenty-first Century*. London: Pegasus.
- Chiapello, E., & Fairclough, N. (2002). Understanding the new management ideology: a transdisciplinary contribution from critical discourse analysis and new sociology of capitalism. *Discourse & Society*, 13(2), 185-208.
- Clancy, J. (1989). *The Invisible Powers: The Language of Business*. Oxford, UK: Lexington Books.
- Clark, P., & Rowlinson, M. (2004). The Treatment of History in Organization Studies: Towards an 'Historic Turn'? *Business History*, 46, 331-352.
- Collection, S. C. P. (2013). Scott Nearing Papers, 1909-1994 Retrieved 20m September, 2013
- Collier, C. (1979). Henry George's System of Political Economy. *History of Political Economy*, 11(1), 64-93.
- Dandridge, T. (1986). Ceremony as an Integration of Work and Play. *Organization Studies*, 7(2), 159-170.
- Debray, R. (1996). *Media Manifestos: On The Technological Transmission of Cultural Forms* (E. Rauth, Trans.). London: Verso.
- Debray, R. (2004). *Transmitting Culture* (E. Rauth, Trans.). New York: Columbia University Press.
- Donaldson, G., & Lorsch, J. (1983). *Decision Making at the Top: The Shaping of Strategic Direction*. New York: Basic Books.
- Donnellon, A., Gray, B., & Bougon, M. (1986). Communication, Meaning, and Organized Action. *Administrative Science Quarterly*, 31, 43-55.
- Durepos, G., & Mills, A. (2012). *ANTI-History: Theorizing the Past, history, and Historiography in Management and Organizational Studies*: Information Age Publishing.
- Durepos, G., Mills, A., & Weatherbee, T. (2012). Theorizing the past: Realism, relativism, relationalism and the reassembly of Weber. *Management & Organizational History*, 7(3), 267-281.
- Eberhard, W. (1968). *The Local Cultures of South and East China*. Leiden, Netherlands: E.J. Brill.
- Edery, D., & Mollick, E. (2010). *Changing the Game: How Video Games are Transforming the Future of Business*. Upper Saddle River, NJ: FT Press.
- Eiss, H. (1994). Preface. In H. Eiss (Ed.), *Images of the Child* (pp. 1-17): Bowling Green University State Popular Press.
- England, R. (2010). Ricardo, gold, and rails: Discovering the Origins of Progress and Poverty. *The American Journal of Economics and Sociology*, 69, 1279-1293.
- Flanagan, M. (2009). *Critical Play: Radical Game Design*. Cambridge, MA: The MIT Press.
- Geertz, C. (1972). Deep Play: Notes on the Balinese Cockfight. *Daedalus*, 101(1, Myth, Symbol, and Culture (Winter)), 1-37.

- George, H. (1912). *Progress and Poverty: An Inquiry into the Cause of Industrial Depressions and of Increase of Want with Increase of Wealth: The Remedy*. New York: Doubleday, Page & Company.
- Goetz, B. (1959). Teaching Production Management. *Academy of Management Journal*, 2, 7-20.
- Goren, P., Federico, C., & Kittilson, M. (2009). Source Cues, Partisan Identities, and Political Value Expression. *American Journal of Political Science*, 53(4), 805-820.
- Gramsci, A. (1968). *The Modern Prince and Other Writings* (L. Marks, Trans.). New York: International Publishers.
- Gramsci, A. (1971). *Selections From The Prison Notebooks* (Q. Hoare & J. Nowell Smith, Trans.). New York: International Publishers.
- Group, D. M. W. D. (2013). *The Village of Arden* Retrieved 20 September, 2013, from <http://www.arden.delaware.gov/>
- Hasbro. (2013). About Monopoly Retrieved 26 September, 2013, from http://www.hasbro.com/monopoly/en_CA/discover/about.cfm
- Heilbroner, N. (1999). *The Worldly Philosophers: The Lives, Times And Ideas Of The Great Economic Thinkers* (7th ed.). New York: Simon & Schuster.
- Hofer, M. (2003). *The Games We Played: The Golden Age of Board Games & Table Games*. New York: Princeton Architectural Press.
- Hoff, K. (1991). Land Taxes, Output Taxes, and Sharecropping: Was Henry George Right? *The World Bank Economic Review*, 5(1), 93-111.
- Huizinga, J. (1949). *Homo-Ludens: A Study of the Play Element in Culture* (R. F. C. Hull, Trans.). London: Routledge and Keagan Paul.
- Jackson, N., & Carter, P. (1993). 'Paradigm Wars': A Response to Hugh Willmott. *Organization Studies*, 14(5), 721-725.
- Jenkins, H., Camper, B., Chisholm, A., Grigsby, N., Klopfer, E., Osterweil, S., . . . Chor Guan, T. (2009). From Serious Games to Serious Gaming. In U. Ritterfeld, M. Cody & P. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (pp. 448-468). New York: Routledge.
- Kane, P. (2004). *The Play Ethic*. London: Pan Books.
- Kapp, K. (2012). *The Gamification of Learning and Instruction: Game based Methods and Strategies for Training and Education*. San Fransisco, CA: Pfeiffer.
- Kark, R. (2011). Games Managers Play: Play as a Form of Leadership Development. *Academy of Management Learning & Education*, 10(3), 507-527.
- Kennedy, J., R. (2004). *Monopoly: The Story Behind the Wolrd's Best-Selling Game*. New York: MJF Books.
- Kieser, A. (1994). Crossroads - Why Organization Theory Needs Historical Analyses - and How This Should Be Performed. *Organization Science*, 5(4), 608-620.
- Kurke, L. (1999). Ancient Greek Board Games and How to Play Them. *Classical Philology*, 94(3), 247-267.
- Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network-Theory*. Oxford: Blackwell Publishing.
- Latour, B., & Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. Princeton, New Jersey: Princeton University Press.

- Law, J. (2009). Actor Network Theory and Material Semiotics. In B. Turner (Ed.), *The New Blackwell Companion to Social Theory* (pp. 141-159). Oxford: Blackwell Publishing.
- Law, J., & Mol, A. (2001). Situating Technoscience: An Inquiry into Spatialities. *Society and Space, 19*, 609-621.
- Leemkuil, H., & de Jong, T. (2012). Adaptive Advice in Learning with Computer-Based Knowledge Management Simulation Game. *Academy of Management Learning & Education, 11*(4), 653-665.
- Lesko, N. (1996). Past, Present and Future Conceptions of Adolescence. *Educational Theory, 46*(4), 453-472.
- Lewin, A., & Weber, W. (1969). Management Game Teams in Education and Organization Research: An Experiment on Risk Taking. *Academy of Management Journal, 12*, 49-58.
- Limon, J. (1983). Western Marxism and Folklore: A Critical Introduction. *Journal of American Folklore, 96*(379), 34-52.
- Lodge, G. (1977). *The New American Ideology*. New York: Alfred A. Knopf.
- Lyotard, J. (1984). *The Postmodern Condition: A Report on Knowledge*. Manchester, UK: Manchester University Press.
- Magie, L. J. (1904). United States Patent No. 748,626: USPTO.
- Magie Phillips, L. J. (1924). United States Patent No. US001509312. USPTO: USPTO.
- Malaby, T. (2007). Beyond Play: A New Approach to Games. *Games and Culture, 2*(2), 95-113.
- Malaby, T. (2009). Anthropology and Play: The Contours of Playful Experience. *New Literary History, 40*(1), 205-208.
- Mannheim, K. (1985). *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York: Harcourt Brace.
- Marx, K., & Engels, F. (1947). *THE German Ideology*. London: International.
- Marx, K., & Engels, F. (2004). Feuerbach: Opposition of the Materialist Outlook: Part One With Selections From Parts Two and Three and Supplementary Texts. In C. Arthur (Ed.), *The German Ideology* (pp. 39-96).
- McGonigal, J. (2011). *Relativity is Broken: Why Games Make Us Better and How They Can Change the World*. London: Penguin Press HC.
- Meier Sorenson, B., & Spoelstra, S. (2012). Play at Work: Continuation, Intervention and Usurpation. *Organization, 19*(1), 81-97.
- Mills, A., Weatherbee, T., & Durepos, G. (2013). Reassembling Weber to Reveal the-past-as-history in Management and Organization Studies. *Organization, forthcoming*.
- Mirowski, P. What Were von Neuman and Morgenstern Trying to Accomplish? *History of Political Economy, 24*(Supplement), 113-147.
- Moss Kanter, R. (1997). *On The Frontiers of Management*. Boston: Harvard Business School Press.
- Murray, H. (1913). *A History of Chess*. London: Oxford university Press.
- Neuhauser, J. (1976). Business Games Have Failed. *The Academy of Management Review, 1*(4), 124-129.
- Orbanes, P. (2006). *Monopoly: The World's Most Famous Game - And How It Got That Way*. Cambridge, MA: Da Capo Press.
- Parush, T. (2008). From "Management Ideology" to "Management Fashion": A Comparative Analysis of Two Key Concepts in the Sociology of Management Knowledge. *International Studies of Management & Organization, 38*(1), 48-70.

- Pisac, A. (2013). Working the Play: How a Card Game Negotiates Perceptions of Work and Productivity. *Narodna Umjetnost*, 50(1. Working Cultures/Cultures of Work Special Issue of), 182-201.
- Postman, N. (1985). The Disappearance of Childhood. *Childhood Education*, 61(4), 286-293.
- Prasad, A. (2013). Playing the game and trying not to lose myself: A doctoral student's perspective on the insitutional pressure for research output. *Organization*, 0(0), 1-13.
- Raelin, J. (2008). Refereeing the Game of Peer Review. *Academy of Management Learning & Education*, 7, 124-129.
- Reall, M., Bailey, J., & Stoll, S. (1998). Moral Reasoning "On Hold" During a Competitive Game. *Journal of Business Ethics*, 17, 1205-1210.
- Reeves, B., & Read, J. (2009). *Total Engagement: Using Games and Virtual Worlds to Change the Way Work and Businesses Compete*. Boston, MA: Harvard Business School Press.
- Rhodes, C., & Parker, M. (2008). Images of Organizing in Popular Culture. *Organization*, 15(5), 627-637.
- Rindova, V., Becerra, M., & Contardo, I. (2004). Enacting Competitive Wars: Competitive Activity, Languange Games, and Market Consequences. *Academy of Management Review*, 29(4), 670-686.
- Ritterfield, U., Cody, M., & Vorderer, P. (2009). *Serious Games: Mechanisms and Effects*. New York: Routledge.
- School, T. W. (2013). A Radical Who Laid the Groundwork for the Tenure System. *Wharton Alumni Magazine*, Spring 2007(20 September). Retrieved from <http://www.wharton.upenn.edu/125anniversaryissue/nearing.html>
- Scrieber, A. (1958). THE THeory and Application of the Management Game Approach to Teaching Business Policy. *Academy of Management Journal*, 1, 51-57.
- Shoup, D. (2004). The Ideal Source of Local Public Revenue. *Regional Science and Urban Economics*, 34, 753-784.
- Starbuck, W. (1982). Congealing Oil: Inventing Ideologies to Justify Acting Ideologies Out. *Journal of Management Studies*, 19(1), 3-27.
- Sugarman, S. (1994). Children on Board: Images From Candy Lands. In H. Eiss (Ed.), *Images of the Child* (pp. 323-333): Bowling Green University State Popular Press.
- Sullivan, T. (2011). How Games Could Save the World. *Harvard Business Review*, 89(6), 134-135.
- Swan, A. (1950-1952). Operations Research in Business and Industry. *Operational Research Quarterly*, 2(4), 67-68.
- Taylor, F. W. (1911). *The Principles of Scientific Management*. New York: Harper.
- University, C. (2013). Columbians Ahead of Their Time Retrieved 20 September, 2013, from http://c250.columbia.edu/c250_celebrates/remarkable_columbians/rexford_tugwell.html
- Vandenbergh, F. (2007). Regis Debray and Mediation Studies, or How Does an Idea become a Material Force. *Thesis Eleven*, 89(1), 23-42.
- von Neuman, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behaviour*. Princeton: Princetown University Press.
- Weatherbee, T., Durepos, G., Mills, A. J., & Helms Mills, J. (2012). Theorizing the Past: Critical engagements. *Management & Organizational History*, 7(3), 193-202. doi: 10.1177/1744935912444358

- Westphal, J. (1998). Board Games: How CEOs Adapt to Increases in Structural Board Independence From Management. *Administrative Science Quarterly*, 43, 511-537.
- Whittaker, H. (2004). Board Games and Funerary Symbolism in Greek and Roman Contexts. *Nordlit*. Retrieved from septentrio.uit.no/index.php
- Whittaker, H. (2006). Game-Boards and Gaming-Pieces in the Northern European Iron Age. *Nordlit*, 20. Retrieved from septentrio.uit.no/index.php
- Willmott, H. (1993). Breaking the Paradigm Mentality. *Organization Studies*, 14(5), 681-719.
- Wittgenstein, L. (1965). *Philosophical Investigations*. New York: The Macmillan Company.
- Wolfe, J. (1973). *The Comparative Learning Effects of a Management Game vs. Casework in the Teaching of Business Policy*. Paper presented at the Academy of Management.
- Zald, M. (1993). Organization Studies as a Scientific and Humanistic Enterprise: Toward a Reconceptualization of the Foundations of the Field. *Organization Science*, 4(4), 513-528.
- Zald, M. (1996). More Fragmentation? Unfinished Business in Linking the Social Sciences and the Humanities. *Administrative Science Quarterly*, 41(2).
- Zichermann, G., & Cunningham, C. (2011). *Gamification by Design*. Sebastapol, CA: O'Reilly Media Inc.

DAMNANT QUODNON INTELLIGENT?
TRINITY WESTERN UNIVERSITY'S SCHOOL OF LAW¹

As the 21st Century progresses and Canadian society becomes more diverse, organizational challenges arise when conflicting worldviews collide. The case study *Damnant Quodnon Intelligent?* explores Trinity Western University's proposal to establish a School of Law and the resistance it faces, principally because of organizational policies that flow from the university's identity as a private, Christian post-secondary institution. The case seeks to invite discussion regarding organizational / societal ethical behavior, the management of dichotomous worldviews, the balancing of competing rights, and the salience of diverse stakeholder perspectives.

A Controversial Proposal

On June 18, 2012, Trinity Western University (TWU) announced that it had submitted a proposal to the Ministry of Advanced Education, Innovation and Technology (MAEIT) and the Federation of Law Societies of Canada (the Federation)² to establish a School of Law at its Langley, BC campus. If approved, TWU planned to begin classes in September of 2015. Opposition to TWU's proposal was immediate, with many detractors taking exception to the school's mandate to follow the teachings of the Bible. Some – including the Canadian Council of Law Deans – suggested that such a school would by definition be discriminatory and would not produce lawyers that would be able to be objective in conserving the law. Others opposed the proposal on the grounds that a number of law schools already exist in British Columbia.

¹ This case is based on historical events and not intended to illustrate either effective or ineffective decision-making. The legal term employed in the title (*damnant quodnon intelligent*) is translated as “they condemn what they do not understand.”

² Every lawyer in Canada, and notary in Quebec, is required to be a member of a provincial law society and to follow the rules that govern the province or territory in which they practice. The Federation is mandated to represent the public interest and to ensure that each lawyer or notary meets the high standards of competence and professional conduct.

The proposal enjoyed the backing of both the University's Senate and Board of Governors³, however as opposition mounted the Board was faced with the prospect that the endeavor might not be approved. While the Board was confident that the proposal met all of the criteria required to offer a law degree, there was concern that negative press might have an adverse effect on the school and its students. And what would be the impact on the institution if the proposal was not approved? Could the proposal legitimately be rejected because of TWU's religious beliefs?

Trinity Western University

TWU's founders established "Trinity Junior College" (TJC) in 1962 with a unique vision for Christian liberal arts and sciences education in Canada. While its first graduating class had only four students, by 1982 the university had an enrolment of over 80. In 1985 TJC was granted full membership in the Association of Universities and Colleges of Canada (AUCC),⁴ leading the BC legislature to recognize the school as a university under its new name – Trinity Western University.

TWU was a comprehensive liberal arts, sciences, and professional studies university committed to continually build upon its "Educational Vitality," which included "educating, transforming, and impacting the world through the wisdom and enterprise of its students, alumni/ae, faculty, and staff" (Exhibit 1). The entire university was based on the Christian faith and "emphasize[d] excellence in academic scholarship, while always keeping in mind how the knowledge and understanding we gain through study can serve the world's deepest needs." It was organized into six schools (Business, Nursing, Education, Human Kinetics, Arts, Media and Culture, and Graduate Studies) and two faculties (Humanities and Social Sciences, and Natural and Applied Sciences), and offered 42 undergraduate and 16 graduate degree programs.

A Proposed School of Law

The development of a law school was seen as the next step in TWU's objectives for the development of the university. Specifically, the law school would aim to develop "highly competent, professional and ethical graduates who will no doubt distinguish themselves in the legal community and beyond [as] a natural and appropriate advancement of the TWU mission." Its mission would focus on leadership development and would work toward the achievement of specific ends statements:

- In the pursuit of missional ends globally, Trinity Western University will bring hope, healing, compassion and reconciliation to the world through Christ-like dedication, service and engagement of its students, alumni, faculty and staff.
- Institutionally, the university will be a world-class centre for vital research, exemplary scholarship, and excellence in comprehensive adaptive, integrated learning.

³ TWU's Senate approved the proposal on April 3, 2012, and the Board of Governors signed its approval on April 29, 2012.

⁴ The AUCC represents 97 public and private not-for-profit Canadian universities and is involved in the development of public policy, the administration of leadership seminars for university leaders, and the management of scholarships and international programs.

- Individually, the university will assist students in becoming those who can change the world by their competence, character, commitment and calling.⁵

Specifically the proposal was to offer a Juris Doctor (JD) program⁶ “with a view to excellence in leadership development, professionalism and public service.” The JD was a graduate level program, and to meet eligibility requirements a prospective student would already possess an undergraduate degree. Building on a common JD curriculum, students would have the “opportunity to specialize in ‘Charities and Social Justice’ or ‘Entrepreneurial’ law, with the possibility of exploring legal approaches to social enterprise and social innovation.” The program would (a) focus on professionalism, practice competence, and high ethical standards; (b) integrate practical assignments; (c) emphasize leadership and character development; and (d) integrate a Christian worldview.

TWU had been planning its offering for many years and felt that this move “[fit] well with the University’s mission to develop Godly leaders for the marketplaces of life.” Janet Epp Buckingham, an associate professor at TWU and Executive Director of the Laurentian Leadership Centre in Ottawa, had initially approached the university with the idea some 20 years prior. She held a firm belief that there was a need for a law school “that had a focus on Christian principles of justice in society where students would be able to discuss what it means to be a Christian and be a lawyer and to have a really different kind of focus on law.”⁷ The university struck an advisory council that consisted of lawyers and judges, and laid down plans to construct a new building to house the law school when the proposal was approved (a process estimated to take between 6 to 12 months).

One part of the rationale behind starting a new law school was to “increase accessibility to a Canadian legal education, as demand for entry currently outstrips the number of law school places available.” Buckingham believed that there were many students who desired to study law but were not able to get into law school because of the limited number of seats available. Accordingly many Canadian students had to attend law schools in Australia, England, and the United States – so many in fact that one Australian law school had hired Canadian professors and adopted a Canadian curriculum.

TWU had developed partnerships with various agencies, which would “give students unique opportunities to provide needed legal services to the less privileged and [represent] again a natural and appropriate advancement of the TWU mission.” Students would be encouraged to view being a lawyer as a “high calling in a life of service to God and the community.” This perspective fit well with the overall mission of TWU. As Buckingham explained, the Canadian legal system has a Judeo-Christian heritage and the Bible has much to say about the law. “There is a huge amount of the Bible that one can look at in terms of legal principles and justice principles. Clearly, law is something God cares about because it involves the relationships we have in our society.” She submitted that TWU’s law school would be different from the others because “law is a client-focused way of life, and so we want to make sure that TWU’s law school

⁵ From TWU’s proposal.

⁶ Juris Doctor (JD) is also known as Doctor of Jurisprudence. In the past Canadian universities offered a Bachelor of Laws (L.L.B.) degree, which was common in other Commonwealth countries. The United States offered the JD degree, and over time a perception developed that the Canadian L.L.B. did not meet the same level of education as the American designation. In 2001 the decision was made to adopt the JD designation.

⁷ From an interview with Peter Stockland.

focuses on the client and helps law students to develop into client-focused professionals.”⁸ From her perspective, other law schools focused not on the client, but on legal principles, and when students arrived at their articling position⁹ they had to quickly learn the skills necessary to work with clients.

Degree Authorization

The MAEIT was the provincial government agency responsible to oversee the assessment of new degree applications in BC, and in order for TWU to implement a new degree program it was required to apply to the Ministry under the BC Degree Authorization Act. A Degree Quality Assessment Board (QAB) would then be struck to assess the application to ensure that it met “consistent and high-quality criteria.” These criteria, developed in harmony with national standards, included:¹⁰

1. Degree Level Standard
2. Credential Recognition and Nomenclature
3. Curriculum/Program Content
4. Learning Methodologies/Program Delivery
5. Admission and Transfer/Residency
6. Faculty
7. Program Resources
8. Program Consultation
9. Program Review Assessment

TWU submitted its proposal and – as per standard practice – it was posted to the MAEIT website for public viewing. When the QAB completed its assessment the proposal was submitted to the MAEIT for review and final decision, with the results of the assessment, the QAB’s recommendations, and the Ministry decision posted to the website along with the final draft of the proposal.

The Federation of Law Societies of Canada and the Canadian Bar Association

The Federation was the national coordinating body for all of Canada’s provincial law societies. Every lawyer (or notary in Quebec) was required to become a member of the law society in the province or territory in which they practiced, overseen in turn by the Federation who’s mission was to “[act] in the public interest by strengthening Canada’s system of governance of an independent legal profession, reinforcing public confidence in it and making it a leading example of justice systems around the world.”

In order to help strengthen the Canadian legal system, the Federation had, according to its mandate, adopted national requirements for law school graduates (Exhibit 4) which were applied

⁸ From an interview with Peter Stockland.

⁹ An articling position is an apprenticeship required of all students desiring to become lawyers. It is typically completed at a law firm and before the writing of the bar exam.

¹⁰ An expanded description is provided in Exhibit 2.

when proposals for new law schools were submitted to the Federation's Approval Committee.¹¹ If a school met the requirements its graduates would be considered to have met the requirements for entry into their respective bar admissions program.¹² Accordingly TWU had concurrently submitted its proposal to the Approval Committee.

The Canadian Bar Association (CBA) differed from the Federation – while Federation membership was mandatory for all lawyers via their provincial law societies, the CBA was an advocacy organization intended to represent the interests of its membership, which was voluntary. The CBA provided lawyers with such benefits as professional development, advocacy in government policy development and law reform, discount pricing for CBA products, and access to conferences, newsletters, magazines, and online resources.

Approaching the Bar

The process of becoming a lawyer in Canada began with the completion of a law degree. Graduates then applied for admission to the law society of the province or territory in which they wished to practice law. Once accepted additional steps included:

- Completion of an apprenticeship program (articling).
- Attendance of a professional legal training course (the bar admission course).
- Passing of licensing examinations (the bar examinations).

Upon completion of all prerequisites a candidate would then be “called to the bar” and admitted as a member of the provincial law society.

Opposition to the TWU Proposal

The Canadian Council of Law Deans¹³ (the Council) opposed TWU's proposal on the basis of the university's character and values as a religious community. Bill Flanagan, Dean and Professor of Law at Queen's University, wrote a letter on behalf of the Council to the Federation challenging the proposal. The Council considered the principal problem to be that TWU “as an evangelical Christian community, has all students who voluntarily enter the university sign on to a ‘community covenant agreement’.” This agreement (Exhibit 3) reflected TWU's mission, core values, curriculum, and community life and stated explicitly that “according to the Bible, sexual intimacy is reserved for marriage between one man and one woman, and within that marriage bond it is God's intention that it be enjoyed as a means for marital intimacy and procreation.” The Council called on the Federation to investigate whether the covenant was inconsistent with federal and provincial law on the basis that it discriminated against gay, lesbian and bi-sexual students.

¹¹ The Approval Committee was appointed by the Federation, which is comprised of seven members – four of which have regulatory experience in Canadian law societies and three of which are Deans at Canadian law schools.

¹² In general “the bar” is considered to mean the legal profession and / or the qualification process whereby a lawyer is licensed to practice law.

¹³ The Canadian Council of Law Deans consists of the heads of the different law schools and exists to enable executive level communication between Canada's providers of legal education.

Constitutional and civil rights lawyer Clayton Ruby, in an editorial in the Toronto Star, stated, “It is just wrong to have a law school approve discrimination in its own structure. That kind of discrimination, which denies some people the right to equality, is fundamentally inconsistent with law and democracy.”¹⁴

Lawyer Robert Brun wrote to the Federation on behalf of the CBA, calling on it to investigate whether TWU’s proposal met the national standards. From the perspective of the CBA, any new law degree program must “strike a balance between freedom of religion and equality, and give full consideration to its public interest mandate and to the values embodied in the Canadian human rights laws.” He went on to state that “the Federation has a duty to go beyond a strict determination of a proposed law school’s compliance with the national standards. It must assess whether the institution and its program complies with Canadian law, including the protections afforded by the Canadian Charter of Rights and Freedoms and the human rights legislation in BC, and in every province and territory where a proposed law degree may be recognized by the law societies for admission to bar.”

Another letter to the Federation from Amy Sadalauskas, Robert Peterson, and Level Chan, representing the CBA’s Sexual Orientation and Gender Identity Conference and the Equality Committee, raised the concern that “the National Standards’ ethical, constitutional and human rights components” had not been met. They submitted that while TWU operated under the Trinity Western University Act of 1969, which stated that TWU “is recognized as a Christian institution affiliated with the Evangelical Free Church of Canada,” the university should not benefit from a suggested exemption from anti-discrimination legislation in the British Columbia Human Rights Code (BCHRC). They reminded the Federation that while BCHRC (s.41(1)) stated that...

41 (1) If a charitable, philanthropic, educational, fraternal, religious or social organization or corporation that is not operated for profit has as a primary purpose the promotion of the interests and welfare of an identifiable group or class of persons characterized by a physical or mental disability or by a common race, religion, age, sex, marital status, political belief, colour, ancestry or place of origin, that organization or corporation must not be considered to be contravening this Code because it is granting a preference to members of the identifiable group or class of persons;

...and that TWU *did not* – according to the Trinity Western University Act – promote the interests of an identifiable group, and that it had not been given authorization to “grant a preference to members of any particular church or religion.” The authors pointed specifically to Subsection 3(2) of the Act which stated:

(2) The objects of the University shall be to provide for young people of any race, colour, or creed university education in the arts and sciences with an underlying philosophy and viewpoint that is Christian.

The authors submitted that “freedom of religion does not allow one group of individuals to exclude another group of identifiable individuals from access to a public service.”

¹⁴ Ruby’s editorial was published March 1, 2013 in The Star, a Toronto newspaper.

Trinity Western's Defense

The university had travelled this road before. In 1997 the British Columbia College of Teachers had opposed its offering of a teacher training program because of its biblical beliefs, stipulating that the school “violated human-rights laws.” The matter escalated all the way to the Supreme Court of Canada who in 2001 sided with the university because the College of Teachers could not prove that TWU graduates would engage in discriminatory actions in the classroom. Nevertheless an editorial in the National Post proclaimed that much had changed since then:

“The crux of the issue is how the discrimination and institutional environment at TWU impacts the ability of the school to teach law. In order to permit entry into a provincial or territorial law society (as determined by the Federation), the law degree program must meet national standards in its curriculum. Those standards require critical thinking about ethical and legal issues. No person can truly think critically from one pre-determined lens, in this case, a lens mandated by TWU.”¹⁵

The university, however, begged to differ, basing its defense in part upon its claimed exemption from certain prescriptions of national and provincial human rights legislation. The Canadian Charter of Rights and Freedoms guaranteed all citizens four fundamental freedoms: (1) freedom of conscience and religion; (2) freedom of thought, belief, opinion and expression, including freedom of the press and other media of communication; (3) freedom of peaceful assembly; and (4) freedom of association, and stipulated that “every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability.” Canada’s Constitution Act stipulated that each province was responsible for legislation relating to education, specifying that “nothing in any such law shall prejudicially affect any right or privilege with respect to denominational schools which any class of persons have by law in the province at the union.”

TWU however was a private institution and maintained that it was exempt, in part, from BCHRC human rights legislation and to which the Canadian Charter of Rights and Freedoms did not apply. Canada’s Human Rights Code stated that non-profit organizations of a charitable, philanthropic, educational, fraternal, religious or social nature could avail themselves of an exemption for the purpose of “the promotion of the interests and welfare of an identifiable group or class of persons characterized by a physical or mental disability or by a common race, religion, age, sex, marital status, political belief, colour, ancestry or place of origin,” and that the “organization or corporation must not be considered to be contravening this Code because it is granting a preference to members of the identifiable group or class of persons.”

Douglas Todd, a columnist for the Vancouver Sun, explained in an article that the president of the BC Civil Liberties Association (BCCLA), Lindsay Lyster, had denounced the Council for its opposition of TWU’s proposed law school. Lyster had accused “the deans of the country’s secular law schools of wanting to ‘monopolize’ legal education and keep ‘religiously-minded’ people from becoming law students and lawyers.” Todd also quoted Lyster as saying that the Council operated on the premise that “those who are religiously minded should be excluded from all legal education. That would, by extension, include all professors, students and

¹⁵ National Post. (2013, January 24). Counterpoint: Why Trinity Western University Should Not Have a Law School. Retrieved April 6, 2013: <http://fullcomment.nationalpost.com/2013/01/24/counterpoint-why-trinity-western-university-should-not-have-a-law-school/>

eventually lawyers and judges who held to the religious views the CCLD says are repugnant.” Lyster believed that in this case the BCCLA supported TWU on the “principle of religious freedom,” and said “secular universities often imposed restrictions on free thought, including in regards to religious perspectives.”

Other Criticisms

Other critics objected to TWU’s proposal, maintaining that there existed a national shortage of articling positions, a fact that would be keenly obvious in BC, given its three schools of law at the University of British Columbia (Vancouver), the University of Victoria, and Thompson Rivers University (Kamloops). An article in Maclean’s magazine (June 25, 2012) suggested that, “given that there is a national shortage of articling positions for law graduates, the decision to open new programs seems curious.”

Interestingly, one reader’s response was to agree that there was a shortage of articling positions and no shortage of lawyers in most of BC (especially in the Lower Mainland area where the proposed school would be located) but that this might be due to the fact that most law school graduates didn’t move to small communities to practice law, but rather settled in to work in the urban areas in which they had studied.

TWU replied that it would be hiring an articling coordinator to help place students – especially in small, under-served communities. “The whole focus is really going to be on building skills alongside building legal analysis and understanding so that when students graduate from the law school they would be able to go into a small firm already with skills that they can use and apply,” said Janet Buckingham. Reiterating her position she explained that many existing law schools are not teaching skills, thus when students graduate and start at their articling position, “they can’t do some of the basic things like drafting contracts and going to court and representing clients.”¹⁶

Damnant Quodon Intelligent?

By July 2013 the prospect of a delayed decision on the proposal appeared possible. While the criteria of the MAEIT appeared to have been met satisfactorily, the Federation seemed hesitant to make a decision, citing the fact that when the national requirements had first been instituted issues such as the ones being raised by the Council and the CBA had not been anticipated and were as such “outside of the mandate of the Approval Committee” (Exhibit 4). To this end the Federation had struck a Special Advisory Committee to investigate the proposal and concerns with it more fully.

For Trinity Western’s part, the university continued to deal with sporadic negative press and the very real prospect that its plans might be delayed, or worse scuttled by a decision against it. Had the university stepped beyond the boundaries of the law? Were its detractors unreasonable in their criticisms? Was there any room for a faith-based school of law in pluralist, post-modern society? The days ahead would prove telling.

¹⁶ From an interview with Peter Stockland.

EXHIBIT 1
TWU's Vision Statement

The Vision

The vision for Trinity Western University's future builds on its identity and essence as a Christian university of the liberal arts, sciences and professional studies, and on its historical mission to develop people of high competence and exemplary character who distinguish themselves as exceptional leaders in "the marketplaces of life."

Within a context of global challenges and opportunities, we envision Trinity Western University as a people, a place, a provision of God, and a priority:

AS A PEOPLE,

the University will increasingly be characterized by educational vitality, full of life; organized for excellence into communities of learning; concentrated on areas of competence and character; called to care about the well-being of others; and committed to knowledge and understanding that addresses the world's deepest needs and most profound opportunities.

AS A PLACE,

the University will increasingly draw learners and scholars from around the world into its various campus communities. Trinity Western University communities will increasingly provide venues and resources that facilitate innovative research and scholarship, deep reflection on the meaning and value of truth, and highly-engaging dialogue on the implications of truth for the world.

AS GOD'S PROVISION,

the University will increasingly be a wonderful gift to the world, a steward of resources to be cherished and further developed, including: intelligent and committed people; attractive and functional lands and facilities; state-of-the-art technology; generous financial aid through scholarships and bursaries; and endowments that support the infrastructure necessary for educational excellence.

AS A PRIORITY,

the University will increasingly attract the finest students, faculty, and staff; engage visionary investors and partners who enable mission fulfillment; and raise the prominence of the University on a local, provincial, national, and international level in a way that recognizes twu's graduates and scholarly activities as among the finest in the world.

Trinity Western University, as a people of vitality, a place of beauty and capacity, a provision of God's remarkable resources, and an institution of priority, will become one of the world's premier Christian universities.

EXHIBIT 2
Ministry of Advanced Education
Degree Program Criteria

1. **Degree Level Standard:** The institution must demonstrate that the proposed program meets or exceeds the proposed degree level standard.
2. **Credential Recognition and Nomenclature:** The institution must demonstrate that the program's learning outcomes and standards are sufficiently clear and at a level that will facilitate recognition of the credential by other post-secondary institutions, professional and licensing bodies and employers. Where appropriate, the program, courses or curricular elements are designed to facilitate credential recognition by other post-secondary institutions and by employers, both within the province and other jurisdictions.

The name of the degree should convey long-term meaning; the content of a degree program should be consistent with the name; and the reputation of the institution and of post-secondary education in British Columbia should be enhanced by the quality of the offering. Beyond that is the value to graduates of having a professional credential recognized by appropriate licensing and accrediting bodies as the basis for entry to practice.

3. **Curriculum/Program Content:** The management structures and methods of the program are well defined and permit delivery of the quality of education necessary for students to attain the learning outcomes. The institution must demonstrate that the program in both subject matter and learning outcome standards, offers an education of sufficient breadth and rigour to be comparable to similar programs at the proposed degree level offered by recognized provincial, national and international post-secondary institutions. The curriculum must be current and reflect the state of knowledge in the field, or fields in the case of interdisciplinary and multidisciplinary programs.
4. **Learning Methodologies/Program Delivery:** Learning methodologies are the methods of delivery that will be used to achieve the desired learning outcomes at the degree level standard and at an acceptable level of quality. The institution must demonstrate that it has the expertise and resources to support the proposed method of delivery and ensure its effectiveness. In some cases, it may be more appropriate for the institution to demonstrate that it has a realistic plan to put the necessary expertise and resources in place.
5. **Admission and Transfer/Residency:** The institution should demonstrate that the program is designed to provide flexible admission and transfer arrangements. Where appropriate, the program, courses or curricular elements are designed to facilitate credit transfer by other post-secondary institutions both within the province and other jurisdictions.
6. **Faculty:** The institution must demonstrate that it has the human resources necessary to develop and deliver a quality degree program. In some cases, it may not be feasible for an institution to hire faculty until it receives program approval. In these cases, the institution should provide the specific faculty selection criteria that will be used to ensure new faculty hires have the necessary qualifications for the degree level being offered and program being proposed.
7. **Program Resources:** The institution must demonstrate that it has the physical, learning, and information resources (both start-up and development) needed to assure a program of acceptable quality. These includes facilities, equipment, library resources, laboratories, computing facilities, shops, specialized equipment, etc., and cooperative work placements where this is a component of the program. In some cases, an institution may not be able to ensure resources are in place until after it receives program approval. In these cases, the institution may bring forward a proposal based on a realistic plan for putting the appropriate resources in place as an alternative to demonstrating that all resources are in place.

8. **Program Consultation:** The institution must demonstrate that it has consulted appropriate individuals and organizations in the development of the program proposal.
9. **Program Review Assessment:** In order to ensure the ongoing currency of the program and the quality of its learning outcomes, the institution must show evidence that a program review and assessment procedure is in place.

EXHIBIT 3

Community Covenant Agreement

Our Pledge to One Another

Trinity Western University (TWU) is a Christian university of the liberal arts, sciences and professional studies with a vision for developing people of high competence and exemplary character who distinguish themselves as leaders in the marketplaces of life.

1. The TWU Community Covenant

The University's mission, core values, curriculum and community life are formed by a firm commitment to the person and work of Jesus Christ as declared in the Bible. This identity and allegiance shapes an educational community in which members pursue truth and excellence with grace and diligence, treat people and ideas with charity and respect, think critically and constructively about complex issues, and willingly respond to the world's most profound needs and greatest opportunities.

The University is an interrelated academic community rooted in the evangelical Protestant tradition; it is made up of Christian administrators, faculty and staff who, along with students choosing to study at TWU, covenant together to form a community that strives to live according to biblical precepts, believing that this will optimize the University's capacity to fulfill its mission and achieve its aspirations.

The community covenant is a solemn pledge in which members place themselves under obligations on the part of the institution to its members, the members to the institution, and the members to one another. In making this pledge, members enter into a contractual agreement and a relational bond. By doing so, members accept reciprocal benefits and mutual responsibilities, and strive to achieve respectful and purposeful unity that aims for the advancement of all, recognizing the diversity of viewpoints, life journeys, stages of maturity, and roles within the TWU community. It is vital that each person who accepts the invitation to become a member of the TWU community carefully considers and sincerely embraces this community covenant.

2. Christian Community

The University's acceptance of the Bible as the divinely inspired, authoritative guide for personal and community life is foundational to its affirmation that people flourish and most fully reach their potential when they delight in seeking God's purposes, and when they renounce and resist the things that stand in the way of those purposes being fulfilled. This ongoing God-enabled pursuit of a holy life is an inner transformation that actualizes a life of purpose and eternal significance. Such a distinctly Christian way of living finds its fullest expression in Christian love, which was exemplified fully by Jesus Christ, and is characterized by humility, self-sacrifice, mercy and justice, and mutual submission for the good of others.

This biblical foundation inspires TWU to be a distinctly Christian university in which members and others observe and experience truth, compassion, reconciliation, and hope. TWU envisions itself to be a community where members demonstrate concern for the well-being of others, where rigorous intellectual learning occurs in the context of whole person development, where members give priority to spiritual formation, and where service-oriented citizenship is modeled.

3. Community Life at TWU

The TWU community covenant involves a commitment on the part of all members to embody attitudes and to practice actions identified in the Bible as virtues, and to avoid those portrayed as destructive. Members of the TWU community, therefore, commit themselves to:

- cultivate Christian virtues, such as love, joy, peace, patience, kindness, goodness, faithfulness, gentleness, self-control, compassion, humility, forgiveness, peacemaking, mercy and justice
- live exemplary lives characterized by honesty, civility, truthfulness, generosity and integrity
- communicate in ways that build others up, according to their needs, for the benefit of all
- treat all persons with respect and dignity, and uphold their God-given worth from conception to death

- be responsible citizens both locally and globally who respect authorities, submit to the laws of this country, and contribute to the welfare of creation and society
- observe modesty, purity and appropriate intimacy in all relationships, reserve sexual expressions of intimacy for marriage, and within marriage take every reasonable step to resolve conflict and avoid divorce
- exercise careful judgment in all lifestyle choices, and take responsibility for personal choices and their impact on others
- encourage and support other members of the community in their pursuit of these values and ideals, while extending forgiveness, accountability, restoration, and healing to one another. In keeping with biblical and TWU ideals, community members voluntarily abstain from the following actions:
 - communication that is destructive to TWU community life and inter-personal relationships, including gossip, slander, vulgar/obscene language, and prejudice
 - harassment or any form of verbal or physical intimidation, including hazing
 - lying, cheating, or other forms of dishonesty including plagiarism
 - stealing, misusing or destroying property belonging to others
 - sexual intimacy that violates the sacredness of marriage between a man and a woman
 - the use of materials that are degrading, dehumanizing, exploitive, hateful, or gratuitously violent, including, but not limited to pornography
 - drunkenness, under-age consumption of alcohol, the use or possession of illegal drugs, and the misuse or abuse of substances including prescribed drugs
 - the use or possession of alcohol on campus, or at any TWU sponsored event, and the use of tobacco on campus or at any TWU sponsored event.

4. Areas for Careful Discernment and Sensitivity

A heightened level of discernment and sensitivity is appropriate within a Christian educational community such as TWU. In order to foster the kind of campus atmosphere most conducive to university ends, this covenant both identifies particular Christian standards and recognizes degrees of latitude for individual freedom. True freedom is not the freedom to do as one pleases, but rather empowerment to do what is best. TWU rejects legalisms that mistakenly identify certain cultural practices as biblical imperatives, or that emphasize outward conduct as the measure of genuine Christian maturity apart from inward thoughts and motivations. In all respects, the TWU community expects its members to exercise wise decision-making according to biblical principles, carefully accounting for each individual's capabilities, vulnerabilities, and values, and considering the consequences of those choices to health and character, social relationships, and God's purposes in the world.

TWU is committed to assisting members who desire to face difficulties or overcome the consequences of poor personal choices by providing reasonable care, resources, and environments for safe and meaningful dialogue. TWU reserves the right to question, challenge or discipline any member in response to actions that impact personal or social welfare.

Wise and Sustainable Self-Care

The University is committed to promoting and supporting habits of healthy self-care in all its members, recognizing that each individual's actions can have a cumulative impact on the entire community. TWU encourages its members to pursue and promote: sustainable patterns of sleep, eating, exercise, and preventative health; as well as sustainable rhythms of solitude and community, personal spiritual disciplines, chapel and local church participation, work, study and recreation, service and rest.

Healthy Sexuality

People face significant challenges in practicing biblical sexual health within a highly sexualized culture. A biblical view of sexuality holds that a person's decisions regarding his or her body are physically, spiritually and emotionally inseparable. Such decisions affect a person's ability to live out God's intention for wholeness in relationship to God, to one's (future) spouse, to others in the community, and to oneself. Further, according to the Bible, sexual intimacy is reserved for marriage between one man and one woman, and within that marriage bond it is God's intention that it be enjoyed as a means for marital intimacy and procreation. Honouring and upholding these principles, members of the TWU community strive for purity

of thought and relationship, respectful modesty, personal responsibility for actions taken, and avoidance of contexts where temptation to compromise would be particularly strong.

Drugs, Alcohol and Tobacco

The use of illegal drugs is by definition illicit. The abuse of legal drugs has been shown to be physically and socially destructive, especially in its potential for forming life-destroying addictions. For these reasons, TWU members voluntarily abstain from the use of illegal drugs and the abuse of legal drugs at all times.

The decision whether or not to consume alcohol or use tobacco is more complex. The Bible allows for the enjoyment of alcohol in moderation, but it also strongly warns against drunkenness and addiction, which overpowers wise and reasonable behaviour and hinders personal development. The Bible commends leaders who abstained from, or were not addicted to, alcohol. Alcohol abuse has many long-lasting negative physical, social and academic consequences. The Bible has no direct instructions regarding the use of tobacco, though many biblical principles regarding stewardship of the body offer guidance. Tobacco is clearly hazardous to the health of both users and bystanders. Many people avoid alcohol and/or tobacco as a matter of conscience, personal health, or in response to an addiction. With these concerns in mind, TWU members will exercise careful discretion, sensitivity to others' conscience/principles, moderation, compassion, and mutual responsibility. In addition, TWU strongly discourages participation in events where the primary purpose is the excessive consumption of alcohol.

Entertainment

When considering the myriad of entertainment options available, including print media, television, film, music, video games, the internet, theatre, concerts, social dancing, clubs, sports, recreation, and gambling, TWU expects its members to make personal choices according to biblical priorities, and with careful consideration for the immediate and long-term impact on one's own well-being, the well-being of others, and the well-being of the University. Entertainment choices should be guided by the pursuit of activities that are edifying, beneficial and constructive, and by a preference for those things that are "true, noble, right, pure, lovely, admirable, excellent, and praiseworthy," recognizing that truth and beauty appear in many differing forms, may be disguised, and may be seen in different ways by different people.

5. Commitment and Accountability

This covenant applies to all members of the TWU community, that is, administrators, faculty and staff employed by TWU and its affiliates, and students enrolled at TWU or any affiliate program. Unless specifically stated otherwise, expectations of this covenant apply to both on and off TWU's campus and extension sites. Sincerely embracing every part of this covenant is a requirement for employment. Employees who sign this covenant also commit themselves to abide by TWU Employment Policies. TWU welcomes all students who qualify for admission, recognizing that not all affirm the theological views that are vital to the University's Christian identity. Students sign this covenant with the commitment to abide by the expectations contained within the Community Covenant, and by campus policies published in the Academic Calendar and Student Handbook.

Ensuring that the integrity of the TWU community is upheld may at times involve taking steps to hold one another accountable to the mutual commitments outlined in this covenant. As a covenant community, all members share this responsibility. The University also provides formal accountability procedures to address actions by community members that represent a disregard for this covenant. These procedures and processes are outlined in TWU's Student Handbook and Employment Policies and will be enacted by designated representatives of the University as deemed necessary.

By my agreement below I affirm that:

I have accepted the invitation to be a member of the TWU community with all the mutual benefits and responsibilities that are involved; I understand that by becoming a member of the TWU community I have also become an ambassador of this community and the ideals it represents; I have carefully read and considered TWU's Community Covenant and will join in fulfilling its responsibilities while I am a member of the TWU community.

EXHIBIT 4
Federation of Canadian Law Societies
Approval Committee Mandate

The Approval Committee have the following mandate:

- To determine law school program compliance with the national requirement for the purpose of entry of Canadian common law school graduates to Canadian law society admission programs. This will apply to the programs of established Canadian law schools and those of new Canadian law schools.
- To make any changes, revisions or additions to the annual law school report as it determines necessary, provided the changes, revisions or additions conform to the approved national requirement and reflect the purposes described in this report.
- To make any changes, revisions or additions to the draft reporting timeline set out in **Appendix 4** and any other reporting timelines as it determines necessary to ensure that the compliance process operates in an effective manner.
- To post its final annual reports on the Federation public website and to post information reports on the website, covering, at a minimum, the list of approved law school programs and issues of interest respecting the continuum of legal education.
- To participate in efforts and initiatives to enhance the institutional relationship between law societies and law schools at a national level. This could, for example, include efforts such as promoting a voluntary national collaboration on ethics and professionalism learning that would further enhance teaching, learning and practice in this area.
- To ensure appropriate training for its members.
- To undertake such other activities and make any necessary changes, additions or improvements to its processes as it determines necessary to ensure the effective implementation of the national requirement, provided these reflect the purposes described in this report.

References

- Association of Universities and Colleges of Canada. (n.d.). *What We Do*. Retrieved on April 6, 2013: <http://www.aucc.ca/about-us/what-we-do>
- Benson, I.T. (2007). The Freedom of Conscience and Religion in Canada: Challenges and Opportunities. *Emory International Law Review*. Vol. 21. (pp. 111-165).
- Brun, R.C. (2013, March 18). *Trinity Western University School of Law Proposal*. Letter Canadian Bar Association.
- CampusAccess.com. (n.d.). *Becoming a Lawyer*. Retrieved April 6, 2013: <http://www.campusaccess.com/education/law-becoming.html>.
- Constitution Act, 1867. *Education*. Retrieved April 16, 2013 from: <http://laws-lois.justice.gc.ca/eng/Const/page-4.html#h-21>.
- Constitution Act, 1982. *Canadian Charter of Rights and Freedom*. Retrieved April 16, 2013 from: <http://laws.justice.gc.ca/eng/Const/page-15.html#h-39>.
- Federation of Law Societies of Canada. (n.d.). *About Us*. Retrieved April 16, 2013 from: <http://www.flsc.ca/en/about-us/>.
- Federation of Law Societies of Canada, (August 2011). *Common Law Degree Implementation Committee: Final Report*. Retrieved July 12, 2013 from: <http://www.flsc.ca/en/national-requirement-for-approving-canadian-common-law-degree-programs/>.
- Federation of Law Societies of Canada. (n.d.). *Our Mission*. Retrieved April 16, 2013 from: <http://www.flsc.ca/en/our-mission/>.
- Federation of Law Societies of Canada. (2009). *Task Force on the Canadian common Law Degree*. Retrieved April 16, 2013 from: <http://www.flsc.ca/en/resources/>.
- Federation of Law Societies of Canada. (2009). *National Requirements for Approving Canadian Common Law Degree Programs*. Retrieved April 16, 2013 from: <http://www.flsc.ca/en/national-requirement-for-approving-canadian-common-law-degree-programs/>.
- Green, J. (2013, March 1). Proposed Christian Law School Should be Denied Accreditation. *The Star*. Retrieved 6 April, 2013: http://www.thestar.com/news/canada/2013/03/01/proposed_christian_law_school_should_be_denied_accreditation_clayton_ruby_says.html.
- Gulli, C. (2010, September 16). The J.D. vs. LL.B degree: Why are schools switching to J.D.? What's the difference, anyway? *Macleans*'s. Retrieved April 16, 2013: <http://www2.macleans.ca/2010/09/16/the-letter-of-the-law/>.
- Human Rights Code*, R.S.B.C. 1996, c. 210, s. 41. Retrieved April 20, 2013 from: http://www.bclaws.ca/EPLibraries/bclaws_new/document/ID/freeside/00_96210_01#section41.
- Joint Statement on Preventing and Resolving Ethical Conflicts Involving Health Care Providers and Persons Receiving Care. (1998). The Canadian Medical Association. Retrieved August 5, 2011: http://www.cma.ca/index.php?ci_id=3217&la_id=1

Newman, D. (2013, January 21). Canadian Law Deans Attack Right to Religious Diversity. *The Vancouver Sun*. Retrieved 6 April, 2013: <http://www.vancouversun.com/health/Canadian+deans+attack+right+religious+diversity/7851195/story.html>.

Macleans: On Campus (2012, June 25). *Trinity Western Wants to Open its Own Law School: Private University Would Have B.C.'s Fourth Law Program*. Retrieved April 6, 2013: <http://oncampus.macleans.ca/education/2012/06/25/trinity-western-wants-to-open-its-own-law-school/>.

National Post. (2013, January 24). *Counterpoint: Why Trinity Western University Should Not Have a Law School*. Retrieved April 6, 2013: <http://fullcomment.nationalpost.com/2013/01/24/counterpoint-why-trinity-western-university-should-not-have-a-law-school/>.

Roshelli, K.M. (2009). Religiously Based Discrimination: Striking a Balance Between a Health Care Provider's Right to Religious Freedom and a Woman's Ability to Access Fertility Treatment Without Facing Discrimination. *St. John's Law Review*. Vol. 83:977. (pp. 977-1016).

Ruby, C. (2013, March 1). *Proposed Christian law school should be denied accreditation, Clayton Ruby says*. The Toronto Star. Retrieved April 14, 2013 from: http://www.thestar.com/news/canada/2013/03/01/proposed_christian_law_school_should_be_denied_accreditation_clayton_ruby_says.html.

Sakalauskas, A., Peterson, R., Chan, L. (March 18, 2013). *Trinity Western University School of Law Proposal*. Letter Canadian Bar Association.

Stockland, P. (December 21, 2012). *Interview: Janet Epp Buckingham on a New Law School in Canada*. Retrieved 18 July, 2013 from: <http://www.cardus.ca/audio/3815/>.

The Canadian Bar Association (n.d.). *Value of Membership*. Retrieved July 12, 2013 from: <http://www.cbamembership.org/Value-of-Membership.aspx>.

Todd, D. (2013, January 30). *How Should Christian Law Schools Handle Gays and Lesbians?* Vancouver Sun. Retrieved April 16, 2013 from: <http://blogs.vancouversun.com/2013/01/30/how-should-christian-law-schools-should-handle-homosexuality/>.

Trinity Western University. (2009). *Community Covenant Agreement: Frequently Asked Questions*. (Unpublished). Langley, BC.

Trinity Western University. (2009). *Envision the Century: Strategic Directions*. (Unpublished). Langley, BC.

Trinity Western University. (2012). *Proposed School of Law at TWU*. Retrieved 6 April, 2013 from: <http://www.twu.ca/news/2012/034-school-of-law.html>.

Trinity Western University. (2012). *Trinity Western University Office of the Provost Program Proposal: Juris Doctor*. Retrieved April 14, 2013 from: <https://www.aved.gov.bc.ca/psips/public/report/applicationsUnderReview.faces>.

Trinity Western University v. British Columbia College of Teachers, 2001 SCC 31, [2001] 1 SCR
772

DAMNANT QUODNON INTELLIGENT?
TRINITY WESTERN UNIVERSITY'S SCHOOL OF LAW
TEACHING NOTE TO THE CASE

Case Summary

Trinity Western University has submitted a proposal to establish a School of Law. This proposal has met with opposition primarily because of the nuances of the university's Christian mandate and secondarily because of a perception that an insufficient availability of articling spaces for existing law programs. Real concern has been expressed from certain stakeholder groups that a graduate from a Christian university will not be able to practice law in an objective and non-discriminatory manner.

Objectives of the Case

The case seeks to invite discussion regarding organizational / societal ethical behavior, the management of dichotomous worldviews, the balancing of competing rights, and the salience of diverse stakeholder perspectives.

Basic Pedagogy

Courses: the case is recommended for use in a Business Ethics course, or in that portion of Strategic Management / Business Policy course that deals with organizational and/or societal ethics.

Level: the case is recommended for use at the senior undergraduate level.

Prerequisites / Co-requisites: Business Ethics

Key Issues:

- Organizational / Societal Ethics
- Worldview
- Competing Rights

Research Methodology

This is a research case based upon secondary public sources.

Relevant Theory

- Ethical Frameworks (utilitarianism, deontology, rights, etc.)
- Legal and Human Rights (conflict, reconciliation)

Discussion Questions

1. **Consider Trinity Western University’s application from the perspective of three different ethical frameworks. What would each suggest is the “right” decision for the Federation to make?**
2. **A central issue in the case would appear to be one of competing rights. Trinity Western University is seeking to establish a school of law that is consistent with its institutional beliefs. Primary objections center on the discriminatory practices enshrined in the university’s policies and call for the termination of the planned expansion. Can organizational values (such as Trinity Western’s) be maintained in a pluralist society, or must they be submerged? Discuss your opinion of the conflict, and whether any means exists whereby these two positions might be equitably reconciled.**

Suggested Responses

1. **Consider Trinity Western University’s application from the perspective of three different ethical frameworks. What would each suggest is the “right” decision for the Federation to make?**

While the authors do not intend for the case to have a decision focus per se, positioning the Federation as a tangible actor upon which students can focus may yield some interesting discussion. It is expected that there will be strong opinions on the matter; however requiring students to address the situation from the perspective of discrete ethical frameworks will ideally force them to consider the facts of the case outside of an emotional response.

The authors recognize that there are many different ethical perspectives from which the Federation’s decision might be approached. For the purposes of the Teaching Note they have drawn upon the schema suggested by Beauchamp, Bowie, and Arnold (2009) and present three points of view, seeking to faithfully apply each framework by first outlining the perspective and then applying it to the scenario.

(a) Utilitarianism

Students might find they are most familiar with this framework given its prevalence in many types of business analysis (e.g. cost-benefit, risk assessment, etc.). Broadly defined, utilitarianism holds that the consequences of actions ultimately determine their worth from a moral perspective. From this viewpoint the Federation should seek to maximize the greatest possible positive value for all persons affected by the decision, and should seek to minimize negative value (or “disvalue”) for all persons affected (which put another way suggests that the greatest good should be done for the greatest number of persons).

It is at this point that students may experience difficulty in generating a utilitarian decision given that they must define what outcome is “good” and determine why this outcome is instrumental for the greatest number of stakeholders. Many will define the population as consisting of the legal community (though some may be tempted to argue at the societal level – which may have worth but may be somewhat nebulous given that case data doesn’t really address this stratum). One approach might be to suggest that the CBA is a proxy for many of Canada’s lawyers and thus represents the “majority.” As such the CBA has made it clear that they oppose the university’s proposal and see approval as being damaging to practice in the legal community. From this perspective utility would seem to suggest that the Federation should not approve the

proposal. Questions, however, must be considered. Does the CBA truly represent the legal community in Canada? No – membership is voluntary and while the CBA does play an advocacy role, the case does not reveal that it is a true proxy for the entire community. Does the opposition of the Council have a role to play from a utilitarian perspective? Doubtful – the Council consists of law school deans, who – while playing a key role in the oversight of legal education in Canada – cannot be construed to represent the majority opinion of the legal community.

One debate that is doubtless to arise is whether the “greater good” can be served by a school that includes discriminatory practices in its human resource and student relations policies. A simplistic declaration that “discrimination is wrong” fails to recognize the nuances of this scenario – students should recognize that this is a minority opinion being expressed contra a dominant, widely accepted standard of behavior. In this case utilitarianism might suggest a grounds to not approve the proposal, although some interesting conversation might take place around whether the exclusion of minority opinions in the legal community ultimately serves it well (i.e. a “monopolization” of legal education as suggested in the case by Lyster).

(b) Kantian Ethics (Deontology)

Named for the works of ethicist Immanuel Kant, this framework centers on the ideas that (a) human beings possess inherent moral value and as such should be treated as ends and not means, and that (b) motives are of more importance than the actions they underlie (ergo actors should make the right decisions for the right reasons). The Kantian framework is inherently deontological in that it defines categorical imperatives (rules) that guide decision making.

Deontology is certainly at play in the case – both the MAEIT and the Federation have a set of criteria (rules) against which they are considering TWU’s application. In the MAEIT’s case it would appear that the QAB is satisfied that the university’s application meets the requirements of the categorical imperatives in effect. In the Federation’s case it at least appears that its requirements are being met – ergo, one might conclude that a Kantian perspective would effectively dictate that because the proposal meets the criteria for the establishment of a law school it should thus be approved. Interestingly, the Foundation – because of the opposition that has been expressed regarding the proposal – appears uneasy with the criteria, suggesting that when they had been established a scenario such as this had not been envisioned.

Some interesting discussion might be had around the Kantian prescription that actions based upon political expediency are prudential, but not moral. For example, were the Foundation to move beyond criteria and not approve the university’s application because of the objections of various stakeholders, such action might be construed as being politically motivated and thus – from a Kantian perspective – an immoral (and therefore wrong) decision.

(c) Rights Theories

Rights form the justifying basis of obligations, of which there are two types: *negative obligations*, which require non-interference with the liberty of others, and *positive obligations*, which require action in the provision of benefits or services to others. Broadly, there are two types of rights: *legal rights*, which are derived from membership in a state or group, and *human rights*, which are independent of membership in a state or group and derived from principles of natural justice.

Much can be made of the conflicting rights in the case. First, notwithstanding the objection of the CBA Committee regarding the Trinity Western University Act (1969) (which

appears to be a somewhat tenuous argument), a cursory reading of the legislative provisions in the case suggests that TWU has a legal right to its position – this might suggest a negative obligation to refrain from interference with the university’s proposal in respect of the institution’s liberty to act within its legal rights. Second, it is also clear that human rights provisos guarantee citizens freedom from discrimination – this might suggest a positive obligation to act to ensure the maintenance of this benefit and thus not approve the university’s proposal. Thus, students are presented with a situation of competing rights, and doubtless will struggle not to elevate one right above another (this could lead to an interesting discussion regarding the criteria employed in the assessment of which right is considered more or less important, a topic broached in Question 2).

- 2. A central issue in the case would appear to be one of competing rights. Trinity Western University is seeking to establish a school of law that is consistent with its institutional beliefs. Primary objections center on the discriminatory practices enshrined in the university’s policies and call for the termination of the planned expansion. Can organizational values (such as Trinity Western’s) be maintained in a pluralist society, or must they be submerged? Discuss your opinion of the conflict, and whether any means exists whereby these two positions might be equitably reconciled.**

In a pluralist society like Canada situations of competing rights must be expected. As outlined in Question 1 (c) two rights are in direct conflict – the university’s legal right to discriminate (broadly defined as freedom of religion), and the rights of individuals in society to be free from discriminatory practice in the context of employment, etc. (broadly defined as freedom from discrimination). It is not unusual in such situations for each side to claim that its rights are inviolable and unable to be subordinated, and students themselves will likely struggle not to subordinate one right to another.

Can these rights be balanced in a manner that is equitable to all involved? Given the inflammatory nature of the conflict, it is unlikely that balance can be achieved in a way that is *satisfactory* to both sides, but *fairness* may be possible. In previous work by one of the authors in this area, it has been suggested that re-framing the problem might be useful, encouraging the student to step away from the facts of the case to consider a parallel situation in another discipline. For example, students could be encouraged to consider the challenges regularly faced in the field of medicine whereby practitioners are often challenged by a conflict of values on religious or moral grounds, pitting “an individual’s right to religious freedom against the government’s strong interest in protecting citizens from discrimination... lead[ing] to frustration and tension between patients and providers.” (Roshelli, 2009) (NB: this perspective requires students to equate organizational values to individual values).

Canada’s largest association of physicians – the Canadian Medical Association (CMA) – stipulates that a doctor need not provide a full range of services (for which s/he is qualified) to the public should the provision of certain services run contrary to individual conscience or religious belief.

Health care providers should not be expected or required to participate in procedures that are contrary to their professional judgment or personal moral values or that are contrary to the values or mission of their facility or agency. Health care providers should declare in advance their inability to participate in procedures that are contrary to their professional or moral

values. Health care providers should not be subject to discrimination or reprisal for acting on their beliefs. The exercise of this provision should never put the person receiving care at risk of harm or abandonment... if a health care provider cannot support the decision that prevails as a matter of professional judgment or personal morality, allow him or her to withdraw without reprisal from participation in carrying out the decision, after ensuring that the person receiving care is not at risk of harm or abandonment. (CMA, 1998)

Benson (2007) suggests that clashes of rights/beliefs (such as those for which the CMA has made allowance) are “unavoidable and are incidents of living in free and open societies... expression, including religious expression, often creates heat as well as light and even successful constitutional claimants must come to recognize that they cannot get everything they seek in the public sphere.” Benson calls for a balance to be struck between conflicting parties, in this instance the doctor and the patient – the doctor must explain the rationale for declining to render service and the patient must seek medical provisions elsewhere provided he/she is not subjected to the “risk of harm or abandonment.” Moving beyond the medical frame Benson states that,

The argument that one citizen treats another as a second-class citizen when refusing to perform a service based upon the exercise of a constitutionally protected right (conscience or religion) focuses the inquiry into the conflict of beliefs in the wrong place. The way to solve a collision of rights is through balancing or reconciling, not submitting the rights of one citizen to another. (Benson, 2007)

Could the same approach be taken within the context of the conflict manifest in the case? Approval of the university’s proposal could be construed as recognition of its right to religious freedom (and legitimate discriminatory practice) – thus its legal rights are upheld. Students and faculty facing discrimination may find service and / or employment at other schools (of which there are many) – thus the “risk of harm or abandonment” would not appear to be significant and human rights will be served (albeit by another provider). Will discrimination occur? Yes, but legitimately so under the maintenance of the balancing – both parties will be spared material harms (although there may be psychological harms that, while difficult to assess in the context of the case, may have a material bearing upon the discussion). In theory both rights can be balanced, which Benson would suggest represents the most equitable approach to such dilemmas:

When a citizen, in an appropriate manner, sets out grounds for a conscientious or religious objection, the duty shifts to the State or employer – that is, to other citizens – to create the appropriate conditions for a reasonable accommodation. Some commentators appear to approach this the other way around and focus the inquiry on responses to the claim of religious liberty rather than where it should be focused – the request for conscientious and religious exemption. To focus on the “feelings” of the person at the other end of the conscientious or religious refusal deflects us from the proper task at hand and would, in almost every case, be a rationale to treat the exercise of

a conscientious or religious refusal as irrelevant. A better example of failure to place any weight at all on the scales of the balance can scarcely be imagined. Exemptions should always be used where the conflict has no other manner of being resolved.

To impose... a supposed “public duty” as a “trump” on individual liberty eviscerates the freedom of conscience and religion and gives no balancing at all when a conflict exists... The quid pro quo that citizens, whether minority or majority, give to other citizens for civic membership is an important aspect of the right of dissent and is essential for both freedom of expression and the public dimension of conscience and religious expression and belief. (Benson, 2007)

Epilogue

At time of writing Trinity Western University’s application to establish a school of law continues to be under review.

ASB 2013
Antigonish, Nova Scotia

Robert A. MacDonald
Heather Steeves
Joanne Laplante
Faculty of Business Administration
Crandall University

HOP-PORTUNITY LOST? SPETZ BREWERY CONSIDERS A PALE FUTURE¹

Small enterprises often struggle with determining the profitability of its product offerings. In some circumstances this may be complicated by the feelings / intuitions of the founder of the business. *Hop-portunity Lost?* is a case study that examines challenges faced by Moncton, NB based Spetz Brewery and sister operation Winexpert, and seeks to provide opportunity to evaluate a business from a managerial accounting perspective to determine whether the business is profitable (and should continue) or unprofitable (and should be shut down).

An Ale of a Problem

Nancy Wheeler let out an exasperated sigh. “Another problem with the fridge!” Ever since it had been installed there had been one problem after another.

Nancy took a moment and looked reflectively around the shop. “The on premise beer brewing² business has never been very profitable,” she thought. From its opening in September 2009 Spetz Brewery had not attracted as many customers as expected, and expensive fridge repairs had been eating up any profit that might have been made. Nancy felt she was in a conundrum - *she* believed that the brewing operation should be shut down, but knew that it meant a lot to her father, George Kraemer, who had founded the business. Although he had *officially* retired in December 2011 and left Nancy in charge, it seemed that he still had the final word in all major decisions.

¹ This case is based on historical events and not intended to illustrate either effective or ineffective decision-making.

² “On premise” is a term used to indicate that the customer makes the wine or beer at the retailer’s location.

Spetz Brewery had been established as one division of the numbered corporation 634384 NB Inc. for the purpose of offering on premise beer brewing to the Moncton, NB marketplace. Sister division Winexpert made up the remainder of the enterprise, and focused on the provision of on premise wine making. While the wine making operation had been consistently profitable, as far as Nancy was concerned the beer-side of the business just seemed to be a hassle.

As Nancy once again dialed the number for the refrigerator repair man, she decided to take another look at Spetz's financial situation. "It's time for a decision to be made," she thought.

The Accidental Vintner

George Kraemer was a long time entrepreneur with a background in engineering. After earning his degree he had specialized in sheet metal fabrication – specifically ductwork for residential and commercial applications – and in the 1960's moved to England to bring this technology to the construction industry there. George found a wide open market – central heating was somewhat of a novel concept to the English as many homes still had coal fireplaces in each room. George's company provided construction firms with sheet metal ductwork fabrication until 1981, when he sold the business and returned to Canada to take over his father's sheet metal fabrication enterprise. As the market changed, George changed with it, moving from ductwork into fire place screens, and then old-fashioned cook stoves.

George moved to Moncton in 1999 when he invested in an M&M Meat Shop franchise.³ Nancy and her husband Brad managed the shop while George enjoyed his "retirement." With a need to stay busy he designed and built blocks for stacking golf balls – which he sold to golf courses around North America – all while writing a book entitled "Faith and Foundation" which told the story of his ancestors and their escape from the Napoleonic wars.

In 2004, George opened a new franchise, Wine Not, which sold wine kits and made wine on premise. He had been interested in home winemaking for many years, and while living in England had made his own wine at home.

About a year after opening, the Wine Not franchise was bought out by Winexpert, the world's largest manufacturer of consumer wine making products. This presented George with the option of becoming completely independent, keeping the Wine Not name and working with Winexpert, or becoming an independent dealer for Winexpert. George decided on the latter, and thus the name of his store changed to Winexpert. One of the hallmarks of the business was the fact that it was the first in the province to make wine on premise. While there were competitors in the area, they only sold kits for customers to take home and it would be 2006 before other on premise businesses popped up.

The Fruit of the Vine

³ The M&M Meat Shop was sold in December 2011 at which time George retired and Nancy began working at Winexpert.

Many customers purchase wine kits because of the enjoyment they derive from making their own wine. In the Canadian marketplace there were four types of wine kit available – these were categorized by the type and purity of the juice involved and ranged from freshly crushed grapes to a combination of juice and concentrate. The price of a kit varied depending on the purity of the juice, with a kit that is pure juice being the most expensive. Included in each kit were the additives required to turn juice into wine.

The rationale behind the growth of on premise businesses was simple: wine making required a fair bit of equipment (see Exhibit 1) which many consumers were hesitant to purchase due to cost or space limitations in their homes. Further, many customers appreciated the economy derived from homemade wine: on premise kits yielding 30 bottles of wine ranged in price from \$129 to \$215 resulting in a cost of approximately \$4.30 to \$7.16 per bottle.

Winexpert

Winexpert was not a franchise, but rather a network of independent dealers. This arrangement required that a free online certification program be completed by each dealer, and that a percentage of the business's inventory consist of Winexpert products (dealers could carry other products as desired). There were no annual licensing fees, and a Winexpert regional representative kept in touch with all of the dealers and served as the first point of contact with the company.

George met regularly with his accountant, who was in charge generating financial statements from the information George provided. It was at one of these meetings in 2006 that the accountant recommended that George incorporate, and thus the numbered company 634384 NB Inc. (operating as Winexpert) was born. The business prospered, and by 2008 operations had grown to the point that the space in a strip mall in Moncton's retail district was too small. Given that George and his wife Penny lived in a house directly across the street, they decided to add an extension to their home⁴ and by 2009 Winexpert had moved across the street.

Spetz Brewery

The decision to brew beer was not accidental - George had been thinking for some time that there was a market for good commercial grade, less expensive beer. He decided on the name Spetz Brewery because his ancestor, Johan Thibeault Spetz, had operated a mini brewing business in Ontario until it was shut down under prohibition in the 1930s.

As George researched the idea of opening an on premise beer brewing business he decided that he wanted to can the beer instead of bottling it, which would provide a product without sediment on the bottom. Home-brewed beer was typically carbonated with sugar, which caused sediment to form on the bottom of bottles. George's visits to small microbreweries revealed that most were employing a cold filtering process (keeping it at about -1 degree Celsius

⁴ When George originally purchased the house it was commercially zoned.

for canning) and carbonating with carbon dioxide (CO₂) instead of sugar and thus generating no sediment. Inspired by his research George searched online and found a website that sold manual canning machines that used CO₂ – shortly thereafter he was installing all of the equipment needed to brew and can beer in the space previously occupied by Winexpert, which he renovated to suit Spetz’s needs.⁵ He purchased a commercial fridge and hired one new employee to run the beer business with an existing Winexpert employee while George and Penny operated Winexpert at the new location.

September 2009 witnessed the grand opening of Spetz Brewery. In the hopes of attracting customers George advertised the business in the local newspaper and distributed flyers throughout the city. He expected that many of Winexpert’s customers – who were wine drinkers – would also want to make beer, but as the months passed it seemed that there were not as many as George had hoped.

Tapped Out

Six months later found George disappointed in Spetz’s performance. The business had struggled to turn a profit, and he found himself considering shutting the operation down altogether. Thus it seemed fortuitous when the owner of a Little Caesars Pizza franchise approached him with an offer to buy out his lease and move in. Before long George had lugged all of Spetz’s gear across the street and into a garage that was attached to the back of Winexpert’s space. While production would take place in the garage, customers would be served in the Winexpert retail space.

With the move complete, George redeployed his two employees. One was moved over to Winexpert while the other was tasked with overseeing Spetz, though because beer brewing wasn’t generating enough activity to support full-time engagement this employee also worked the wine side. George also looked for other ways to cut costs. Although he had been in favour of canning, he knew that cans drove up his costs as they had to be shipped from Alberta. Cans also took up a lot of storage space, which George did not have in the garage. After giving it some thought he made the decision to switch to bottles, which were both cheaper and reusable. He sold his canning equipment and purchased a tap for filling bottles.

The Market for Suds

George’s experiences in England had shown him that Europeans drink beer all year long, so when he began on premise beer brewing in Canada he expected demand to be much the same. To his surprise he found the business to be very seasonal – things picked up in February and March as customers would begin making their beer in anticipation of the warmer weather. The busy period ended in June to be followed by a spike in late fall in anticipation of the Christmas season.

Because there were no other businesses using CO₂ in their brewing processes and

⁵ While Winexpert had moved across the street, George continued to lease space in the strip mall.

providing a bottling service George believed Spetz stood out from its competition. There were five other businesses in Moncton that sold beer kits, but only two of these made beer on premise. One bottled beer for its customers, but used sugar to achieve carbonization (with its resultant sedimentation). The other used CO₂ but didn't bottle the beer for the customer opting instead to fill kegs (from which the customer would later dispense their beer).

Yet in spite of George's perceived competitive advantage, he found that many customers did not repeat their business with Spetz, and he wondered if it might not have something to do with the nature of his beer. George and his staff had learned that because Spetz's beer was not pasteurized⁶ it had to be kept cool (below 14°C). He presumed that many customers did not have extra fridge space at home and thus would store their bottles at room temperature – this would cause the yeast to begin to re-ferment which in turn caused the beer to develop a “skunky,” stale taste. Kit-beer had a shelf life of only three to four months, so if a customer wasn't a frequent beer drinker the product would go bad before they had time to consume it, and if a customer had to write-off most of their purchased product, they might be put off the idea of brewing on premise beer again.

The On Premise Brewing Process

The beer brewing process began when a customer walked into the store. Although Winexpert emphasized on premise wine (and beer) making, the staff always asked first-time customers whether they wished to purchase the kit to make at home or on premise, as kits were priced accordingly (Table 1). If the customer decided to make the beer at home, payment would be made for the kit and any supplies needed. These “home brewers” represented about 15% of sales. Alternatively, if the customer desired to make beer or wine on premise then the price increased accordingly to include the cost of the bottles and caps, labels, processing fees, and filtering.

Winexpert Price List					
Name of Beer Kit	Price for Kit	Price for On premise Brewing	Name of Wine Kit	Price for Kit	Price for Making On premise
Barons Blue Label	\$27.95	\$74.95	Island Mist	\$55.00	\$110.95
Barons Green Label	\$32.95	\$79.95	Vintner's Reserve	\$100.00	\$155.95
Festa Premium	\$39.95	\$94.95	World Vineyard	\$120.00	\$175.95
Festa Specialty	\$44.95	\$99.95	International Series	\$130.00	\$185.95
			Eclipse	\$175.00	\$230.95

Table 1. Price List

⁶ Pasteurization involves heating and then cooling a liquid to inhibit microbial growth and slow the spoilage process.

The business fell under the jurisdiction of the New Brunswick Department of Public Safety, which required certain regulations⁷ to be followed for on premise brewing (Exhibit 2). For example, upon purchase of a kit, the customer had to be present when the kit was prepared. The customer had to personally add the yeast to begin the fermentation process (the yeast creates the alcohol and staff were permitted to initiate the process). From there staff would handle the rest of the process until it was time to bottle at which point regulations required the customer to return, bottle the beer, and take it home immediately.

The Cost of Making Beer

Nancy believed that on premise beer making was costing more than it was worth. She had considered the company's 2012 performance across both beer and wine activities and was not content with what she saw: 256 beer kits had been sold for on premise brewing compared to 2,168 wine kits. Further, the fridge since its installation had needed significant repairs – about every four months a repairman would have to be called and the resultant expense had run from \$400 to \$1,400. Denis, who worked on the beer side of operations, believed that the fridge was not installed properly and that it could possibly be fixed with the next maintenance call. Nancy wasn't so sure – the level of sales on the beer side of the operation just didn't seem high enough to warrant this continued expense. And then there was the cost of electricity associated with running the beast. It seemed to her that the power bill rose and fell with the sale of beer kits, and it was not unusual to have only one keg cooling at a time.⁸ Maybe the time had come to eliminate on premise beer brewing and the refrigeration unit in one fell swoop, and simply continue to offer the home brew kit for sale.

Costs associated with brewing the beer were fairly straight forward (Table 2), with wine making costs approximately 75% higher.

Spetz Brewery Costs				
	Barons Blue Label	Barons Green Label	Festa Premium	Festa Specialty
Cost of Kit	\$15.00	\$16.00	\$18.00	\$22.00
Labour	18.00	18.00	18.00	18.00
CO2	4.00	4.00	4.00	4.00
Labels	0.49	0.49	0.49	0.49
Filters	0.87	0.87	0.87	0.87

Table 2. Costs per Kit

⁷ Winexpert's license allowed it to hold alcohol on premise in bulk only and not in bottles, and forbade the sale of alcohol, permitting only the sale of the ingredients necessary to enable the making of alcohol.

⁸ It was possible to store up to 16 kegs at one time in the fridge, though there was typically between one and five kegs cooling at any one time.

Denis was the only employee that worked on the beer making side of the operation. He was paid \$12 an hour and from the time of purchase to bottling each kit took 1.5 hours to complete. Other expenses included utilities, 50% of which was allocated to the beer making operation (located separately from the winemaking operation in the climate-controlled garage which included the fridge), and 50% to the winemaking operation (this included the fermentation room⁹ and the retail store) (Table 3). Nancy considered utilities to be a major contributor to the cost of each beer and wine kit. The beer making operation occupied approximately 1,000 square feet of space while the winemaking operation and retail space occupied approximately 3,000 square feet. When the fridges were operating they generated heat, which required air conditioning to maintain the temperature of the room. The beer making room was kept at a temperature of 22° Celsius, and when beer was first made it was kept at this temperature for 21 days after which it was moved into the fridge at -1° Celsius for four days to cool. The beer was then carbonated with carbon dioxide, filtered, and transferred to a keg. The keg remained in the fridge for another two days in preparation for bottling, and if it got too warm, the beer would foam when poured into the bottles. Wine, on the other hand, had to be maintained at temperature of 22° Celsius for the entire production process.

Utilities Expenses			
2010/2011		2011/2012	
October	524.23	October	458.07
November	525.44	November	447.92
December	658.70	December	577.35
January	667.96	January	766.33
February	544.13	February	623.32
March	635.30	March	623.01
April	620.88	April	571.52
May	591.18	May	469.09
June	617.74	June	496.12
July	559.09	July	450.73
August	474.88	August	427.56
September	482.60	September	366.73

Table 3. Utilities Expense

The Department of Public Safety required that the number of liters of alcohol made in on

⁹ The fermentation room is a climate-controlled room where the wine is kept during processing.

premise operations be recorded and reported (Table 4).¹⁰ Each kit made 23 liters of beer or wine, or 45 bottles of beer and 30 bottles of wine.

Quantity of Wine and Beer Sold		
	Wine (Liters)	Beer (Liters)
2011		
April	5,405	322
May	4,600	667
June	3,910	437
July	2,944	345
August	5,290	483
September	3,910	460
October	5,520	345
November	2,990	621
December	2,323	299
Total	36,892	3,979
2012		
January	4,830	368
February	4,140	483
March	5,290	759
April	4,531	736
May	5,198	644
June	4,255	759
July	3,197	460
August	4,393	368
September	3,542	391
October	4,439	230
November	3,772	483
December	2,277	207
Total	49,864	5,888

Table 4. Kit Sales in Liters

Other Costs

Nancy had given some thought to the impact closing Spetz might have on expenses. She knew it was the major contributor to maintenance charges (Exhibit 3). If Spetz were to be shut down, the garage could be opened up and used as storage for winemaking equipment, bottling supplies and other product inventory. Air conditioning and heating would still have to be used to

¹⁰ Prior to April 2011 amounts were not tracked on a monthly basis since the law only required annual sales amounts.

control the temperature, but the problematic refrigerator would be eliminated. Other expenses would not be impacted by the closure – Nancy concluded that the employee who worked on the beer side of the business would simply shift over to the wine side, given that he worked an average of 20 hours per week and only spent 25% of his time on the beer operation anyway.

Hop-portunity Lost?

Nancy was tired of fixing the chronically broken fridge and watching Spetz's profits go down the drain. She also believed that the electricity expense to run the fridge was too high, which just seemed to add insult to injury. As she hung up the phone after having summoned the repairman, she let out a long sigh. "It's time to take the bull by the horns," she thought. She dug out her accounting records to attempt to make some calculations around the profitability of the beer business. Her intuition told her that it was time for Spetz to die, and to simply carry on with Winexpert, selling beer kits to home brewers. But what would the numbers say?

EXHIBIT 1
Wine Making Equipment

Basic Equipment required for home winemaking	
Equipment	Description
Gallon jug	Glass container used exclusively to prepare the sanitizing solution to clean your equipment.
Primary fermenter	Food-grade plastic pail in which to start the batch which should include a cover.
Large measuring cup (2 quarts or bigger)	Large enough to measure and pour the water for the recipe.
Long-handled plastic (food-grade) spoon	Large enough to mix a 5-gallon batch.
(Photo)Hydrometer	This two-piece set (including the hydrometer and the plastic chamber) allows you to measure the specific gravity (SG) of the wine must. This gives an indication of how much alcohol is in the wine.
Siphon hose (5 feet)	This PVC-grade plastic hose allows you to transfer wine from the primary fermenter to a carboy or from one carboy to the next.
Two glass carboys (19 litre or 5-gallon)	Containers that come to a small opening at the top.
Airlock and rubber bung	This water-filled plastic device set into the rubber stopper allows carbon dioxide gas to escape from the carboy while preventing air from getting in.
Large food-grade plastic funnel	A funnel will make it easy to transfer cleaning agents or any additives you may want to pre-mix with extracted wine.
Wine thief	This long, tubular device is for extracting must and wine samples from a carboy.

EXHIBIT 2
New Brunswick Department of Public Safety Regulation

NEW BRUNSWICK REGULATION 84-265
under the
LIQUOR CONTROL ACT (O.C. 84-936)

UVIN/UBREW LICENCE

62.2(1) A UVin/UBrew licensee shall ensure that a customer visits the UVin/UBrew establishment at least 2 times and carries out the following tasks:

(a) during the first visit, the customer shall

- (i) if applicable, pay for the ingredients for the manufacture of wine or beer;
- (ii) pay for the services for the manufacture of wine or beer at the UVin/UBrew establishment; and
- (iii) mix all the ingredients necessary to start the fermentation process to manufacture the wine or beer; and

(b) during the second visit, the customer shall

- (i) if the bottles to be used for bottling the wine or beer bear the labels of wineries or breweries, remove the labels from the bottles;
- (ii) bottle the wine or beer and cork or cap the bottles;
- (iii) package the bottles of wine or beer; and
- (iv) remove the wine or beer from the UVin/UBrew establishment.

62.2(2) A customer may be accompanied by a person who assists the customer in performing the tasks mentioned in subsection (1) as long as that person is not the UVin/UBrew licensee.

62.2(3) If the customer is physically incapable of performing the tasks mentioned in subsection (1), the UVin/UBrew licensee may perform the tasks on the customer's behalf in the customer's presence.

62.3 The customer shall carry or convey his or her wine or beer in unopened bottles to his or her residence and any person who is not prohibited from consuming liquor may consume the wine or beer at the customer's residence.

62.4(1) A UVin/UBrew licensee shall provide the customer with an invoice setting out the following information:

- (a) the customer's name, address and telephone number;
- (b) whether the customer is manufacturing wine or beer and the quantity of the wine or beer;
- (c) if applicable, the ingredients provided to the customer and the price charged for the ingredients;
- (d) the services provided to the customer and the price charged for the services;
- (e) the date on which the customer begins manufacturing wine or beer;
- (f) the amount of payment received from the customer;
- (g) the name, address and telephone number of the UVin/UBrew establishment; and
- (h) a statement indicating that the wine or beer is manufactured for consumption exclusively at the customer's residence.

EXHIBIT 3
634384 NB Inc.
For Year Ended 2011 and 2012

634384 NB Inc.		
Income Statement		
For Year Ended September 30		
	2011	2012
Sales	\$397,667	\$402,775
Expenses		
Cost of Goods Sold		
Purchases	\$164,729	\$167,083
Freight	1,011	1,125
Total Cost of Goods Sold	\$165,740	\$168,208
Payroll		
Wages & Levies	\$51,075	\$194,680
Total Payroll Expense	\$51,075	\$194,680
General and Administrative		
Accounting & Legal	\$5,200	\$410
Advertising & Promotions	454	9,454
Courier & Postage	80	149
Credit Card Charges	7,418	7,354
Insurance	(34)	2,444
Interest Charges	-	43
Bank charges	480	435
Office Supplies	672	770
Cleaning Expense	3,578	1,802
Computer Support	485	240
Donations	650	108
Miscellaneous	4,596	4,527
Professional Fees	4,300	5,450
Uniforms	14	-
Rent	42,000	42,000
Repair & Maintenance	3,413	5,463
Telephone	1,708	2,259
Travel & Entertainment	(26)	758
Utilities	6,902	6,278
Total General and Administrative	\$81,891	\$89,944
Total Expenses	\$298,706	\$452,832
Net Income	\$98,961	(\$50,057)

Bibliography

Matthews, E. (Spring 2000). *Your First Wine from a Kit*. WineMaker Magazine. Retrieved from: <http://www.winemakermag.com/new/kit>.

New Brunswick Department of Public Safety. (October 29, 1984). *New Brunswick Regulation 84-265 under the Liquor Control Act (O.C. 84-936)*. Retrieved from: <http://www.gnb.ca/0062/PDF-regs/84-265.pdf>.

**HOP-PORTUNITY LOST?
SPETZ BREWERY CONSIDERS A PALE FUTURE
TEACHING NOTE TO THE CASE**

Case Summary

Nancy Wheeler, manager of Winexpert, has to make a decision on whether or not to shut down the on premise beer brewing side of the business. While Winexpert was established as on premise wine brewing business, in 2009 company owner George Kraemer began offering on premise beer brewing with the belief that there was a market for this product. When George retired in 2011 his daughter, Nancy, took over the day-to-day management of the business and as she examines Nancy examines the books she is seeking to determine the veracity of her belief that the brewing business is unprofitable.

Objective of the Case

The case seeks to provide students with the opportunity to evaluate a business from a managerial accounting perspective to determine whether the business is profitable (and should continue) or unprofitable (and should be shut down).

Basic Pedagogy

Courses: this case is recommended for use in Managerial / Cost Accounting classes.

Level: this case is recommended for students in an undergraduate business program.

Prerequisites: students should have completed or be enrolled in a Managerial or Cost Accounting course.

Key Issues

- Cost behavior
- Evaluation of closing a division

Research Methodology

Primary research for this case included interviews with the case principal. Secondary research included the consultation of online resources regarding the “UBrew” industry.

Relevant Theory

- Least-Squares Regression Method
- Cost-Volume-Profit Analysis: Multiple Products
- Decision Making: Relevant costs

Discussion Questions

1. **Is the manager's assumption of a correlation between the number of beer kits sold and the rise and fall of the utilities bill correct? Identify and assess any other assumptions made by the manager that could impact this calculation.**
2. **Has Winexpert achieved break-even in the on premise beer brewing operation?**
3. **Should Winexpert drop the on premise brewing operation and just sell kits to home brewers?**

Suggested Responses

1. **Is the manager's assumption of a correlation between the number of beer kits sold and the rise and fall of the utilities bill correct? Identify and assess any other assumptions made by the manager that could impact this calculation.**

Students may choose one of a number of techniques to determine whether a correlation exists and while visual fit or high-low methods commonly taught in introductory textbooks are both simple and easy to apply, neither one provides a measure of how well units processed explains fluctuations in utilities cost. Simple linear regression analysis will provide this measure of *goodness of fit* by calculating R^2 and would therefore be the superior choice in this application.

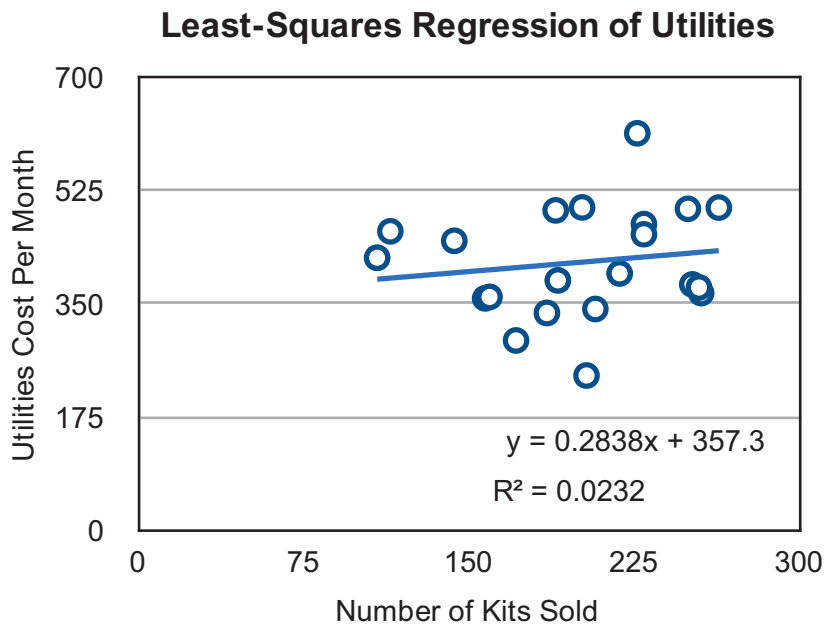
The case offers several real life complications for the student to address in applying a regression concerning data collection and the selection of the simple regression model. The first is the recognition that the information available by month for utilities spans 24 months whereas the information for sales exits for a shorter timeframe. The instructor has a lovely opportunity here to push students to examine whether the timeframe available is sufficient to draw long-term conclusions regarding the cost behavior of utilities expense and to address the limitations of using past cost behavior to predict future cost behavior, establishing the difference between the historic reporting purpose of accounting information and the need to use this information for predictive purposes in decision making (see question 3 for an expanded discussion of relevance).

A second issue surrounding data collection involves matching time periods. By reviewing the financial information students should understand that the expense numbers provided, including utilities, are prepared on a cash basis not on a usage basis, and while this is likely appropriate given the short time frames involved, the choice to use this imperfectly matched cost – sales relationship should be a conscious one, not one made by oversight.

Finally the selection of a simple linear regression analysis vs. a more complex multiple linear regression should be examined. The student should question whether the utilities bill is more likely driven by a single variable (number of units processed) or multiple variables (ambient temperature, building volume, number of units processed, etc.) If the model selected is the wrong model, then its ability to predict futures costs is compromised. Students will likely need to be pushed to consider the limitations of the model and its relevance for this case. Overall, they should conclude that when examined against the information available and the sophistication of the manager, the simple linear regression is adequate for the purpose.

An interesting reporting reality is that the reports made to government regulators are made in liters, whereas the reporting within the business for decision making purposes is prepared by “kit”. In order to maintain consistency within the calculations in Questions 1 and 2, a well thought out answer will first examine total liters of beer and wine sold and convert these numbers into the number of kits sold using the 23 liters/kit conversion factor provided in the case (calculations provided in TN Exhibit 1). With these adjustments made, the least-squares regression calculation can be performed to estimate the cost behavior of the utilities expense (TN Exhibit 2).

The graph following plots the least-squares regression line for the utilities expense. The fixed component of the expense is \$357.30 per month and the variable component per kit is \$0.2838.



To evaluate whether this regression line is economically plausible, the goodness of fit (R^2) is assessed. The goodness of fit has been calculated to be 2.3%, indicating that this regression line is not a good fit for the data and that it would not be a good predictor of future cost behavior. Therefore, this analysis does not demonstrate a strong correlation between the

number of kits processed on premise and the rise and fall of the utilities bill.

2. Has Winexpert achieved break-even in the on premise beer brewing operation?

The break-even point is the point at which total revenues and total costs equals 0. The calculation of the break-even point – considering that there are two products – is calculated using the sales mix percentage of the two products:

CVP Analysis: Multiple Products (FY 2012)					
Product	Kit Price	Unit VC	Unit CM	Units Sold*	Sales Mix
Wine Kit	172	40.44	131.56	2,183	88.96%
Beer Kit	87.45	23.11	64.34	271	11.04%
Total				2,454	

An average of the kit costs was used in this calculation. Unit variable costs for on premise wine making is 75% higher than the beer brewing variable costs ($23.11 + 75\% = \$40.44$). The sales mix is then used to calculate the weighted-average unit contribution margin, which is as follows:

$$\begin{aligned}\text{Weighted-average CM} &= (131.56 \times 88.96\%) + (64.34 \times 11.04\%) \\ &= \$124.1\end{aligned}$$

The break-even point is as follows:

$$\begin{aligned}\text{Break-even point} &= \frac{\text{Fixed Costs}}{\text{Weighted-average CM}} \\ &= \frac{284,624}{124.14} \\ &= 2,293 \text{ kits}\end{aligned}$$

In light of the sales mix, the break-even point for each product is calculated as follows:

$$\begin{aligned}\text{Wine kits: } & 2,293 \times 88.96\% = 2,040 \text{ kits} \\ \text{Beer kits: } & 2,293 \times 11.04\% = 253 \text{ kits}\end{aligned}$$

This analysis shows that Spetz exceeded the break-even point for both products in 2012, although this calculation is only valid if the sales mix doesn't change. More importantly, the

manager (and the student) can use this information in the final question to inform the decision of whether beer kits should be dropped from the product line.

NB: In calculating the break-even the number of units sold was calculated from October 2011 to September 2012 - the fiscal year of the company.

3. Should Winexpert drop the on premise brewing operation and just sell kits to home brewers?

The decision to drop the on premise brewing service requires the student to determine which information is relevant. Students struggle with this distinction often commenting that “every question is different” or “we didn’t include that last time,” failing to grasp the required conceptual understanding of relevance and trying to memorize a format used successfully in the past. This is especially true for accounting students who have completed introductory financial accounting courses which tend to be driven by considerations of format, consistency and repetition.

A conceptual understanding of decision relevant information has two components: information is relevant when it is future oriented (and therefore not relevant when it references past spending, actions or decisions), and it differs for each alternative (and therefore not relevant when the future action is common among alternatives). While many students are able to recall this definition of relevance, the “messiness” of a case provides the perfect opportunity to apply this overtly conceptual definition.

Quantitative Analysis:

The case explicitly states the problem: should the on premise brewing be eliminated? The student will assume either a “Keep” or “Drop” state and craft the quantitative analysis from that viewpoint. The sample analysis following assumes “Drop” in keeping with the manager’s viewpoint and therefore if the bottom line renders a positive number the on premise brewing operation will be dropped; if the outcome is negative the operation will be kept. In terms of determining the relevance of information to use in the analysis, the information can be conveniently understood in two categories: contribution margin (CM) and fixed cost.

Contribution margin (i.e. selling price and variable costs) is typically relevant. If Winexpert “Keeps” the operation, CM will continue in the future; if it “Drops” the operation the CM will cease therefore meeting both the *future oriented* and *varies under alternatives* criteria. The calculation of CM will create the same problems and required assumptions as the student faced in Question 2. With multiple products of varying selling price and no product line sales information, a simplifying assumption regarding CM must be made - for example using a simple average CM.

Average Contribution Margin per Unit		
Average selling price		\$87.45
Average variable costs:		
Cost of kit	\$17.75	
CO ₂	4.00	
Labels	.49	
Filters	.87	
Utilities	.28	
Total variable costs		23.39
Average CM per unit		\$64.06

This calculation was made using information provided in Table 2 of the case and the variable per kit utilities cost calculated in Question 1. Two areas of complication should be noted. First, only the variable portion of the utilities is included in the calculation. The fixed portion, because it is volume driven will not change by eliminating the on premise brewing. Therefore it is not included in the analysis. Second, direct labour has also been excluded from the calculation – while it is appropriate for the purpose of full costing the product for pricing, it is not eliminated if on premise beer brewing is dropped. The employee will simply be shifted to other duties on the wine side of the business where he also currently works. Therefore the labour cost which may appear “variable” is excluded. The student may argue the point that this cost is in fact eliminated from the beer production costs – which is true – but the manager must think holistically in terms of the impact on the business in total. A transfer of costs from one division to another is not a net cost savings to the business.

Calculating total contribution margin will require the student to make sales volume assumptions for the future. Using TN Exhibit 1, annual kit sales are quickly calculated as 245 units. The reasonableness of this number is borne out by its consistency since the start of the operation. Other sales volumes could be used, but this one is selected as a conservative alternative, assuming no future growth. Using this average with the information calculated above provides a total annual CM from on premise brewing of \$15,694.70, which would be given up by ceasing operations.

Ultimately, if the business is to benefit from dropping on premise brewing there must be cost savings greater than the sacrificed CM. Since variable costs have been included in the CM calculation, the student should review any fixed costs that could be eliminated. By definition fixed costs are not impacted by changes in volume unless capacity is eliminated, but that is not the case here since capacity will either revert to wine production or sit idle, neither of which will create cost savings for the business. However, the manager suggests that the cost of repairing the refrigeration unit could be eliminated and is creating a loss in this product line and it is in fact this factor that has contributed to Nancy’s desire to drop the product. Interestingly this view is not shared by her employee who suggests that “the fridge was not installed properly and that it was possible that it could be fixed with the next maintenance call.”

Once again the student will have to exercise judgment when including this number. Any

incremental cost savings associated with no longer running and repairing the refrigerator should be included in the analysis, but estimating this number under conflicting opinions is not obvious. A reasonable alternative is to choose to be conservative in spite of what Denis says and include an amount comparable to that which has been spent in the past. Nancy suggests that “about every four months a repairman would have to be called and the resultant expense had run from \$400 to \$1,400” or up to \$4,200 per year. This is confirmed by reviewing the financial information that presents repairs and maintenance as \$3,400 and \$5,500 (2011, 2012 respectively) for the entire organization. Using the \$4,200 will render the following:

Impact of Eliminating On premise Beer Brewing	
	“Drop”
Contribution margin per unit	\$64.06
Projected annual sales units	245 units
Total annual CM (lost)	(\$15,694.70)
Annual Costs Saved:	
Repairs and Maintenance	4,200.00
Net Annual Gain (Loss) Created by Dropping On premise Brewing	(\$11,497.70)

It is clear from the quantitative analysis that dropping the on premise brewing operation will conservatively cost the company \$11,500 annually. Given such a clear financial mandate to continue the operation, contrary to the manager’s view, students should consider any qualitative factors that are influencing Nancy’s decision.

A final note in the quantitative analysis would recognize that on premise beer brewing at 256 units is currently operating at 38% above its break-even point of 185 units as calculated in Question 2 and – as mentioned in the narrative – this has been achieved with very little in the way of promotional spending. This level of safety margin provides a positive return on this operation with very low risk of loss.

Qualitative Analysis:

Can the nuisance factor be sufficiently quantified? On page 6 there is an interesting statement: “Nancy believed that on premise beer making was costing more *than it was worth*” (emphasis added). Since it is clearly “worth it” financially to continue the operation, some speculation as to qualitative factors that are influencing her is warranted. Questions to consider include:

- Can they grow the beer market?
- Is beer interfering in the potential growth of the wine business? (Analysis can demonstrate that there is sufficient capacity to do both. If capacity was constrained, wine is the better choice since it has a higher per kit CM. With no constraints they should produce both to the extent that customers desire both.)

- Is wine the better product in some (unidentified) intangible way?
- Beer is George's "baby". What issues exist between absentee father / owner and daughter/manager?

Epilogue

Winexpert ultimately decided to discontinue making beer on premise, although it would continue to sell beer kits. As at the time of writing, the business had discontinued stocking the Festa brand (because it wasn't a Winexpert brand) even though it was more popular among who bought kits to brew at home. The wine side of the business remains profitable and continues to grow.

TEACHING NOTE EXHIBIT 1

Quantity of Beer and Wine Sold

Quantity of Beer and Wine Sold						
2011						
	Wine (Liters)	Wine Kits	Beer (Liters)	Beer Kits	Total Liters	Total Kits
April	5,405	235	322	14	5,727	249
May	4,600	200	667	29	5,267	229
June	3,910	170	437	19	4,347	189
July	2,944	128	345	15	3,289	143
August	5,290	230	483	21	5,773	251
September	3,910	170	460	20	4,370	190
October	5,520	240	345	15	5,865	255
November	2,990	130	621	27	3,611	157
December	2,323	101	299	13	2,622	114
Sub-Total	36,892	1,604	3,979	173	40,871	1,777
2012						
January	4,830	210	368	16	5,198	226
February	4,140	180	483	21	4,623	201
March	5,290	230	759	33	6,049	263
April	4,531	197	736	32	5,267	229
May	5,198	226	644	28	5,842	254
June	4,255	185	759	33	5,014	218
July	3,197	139	460	20	3,657	159
August	4,393	191	368	16	4,761	207
September	3,542	154	391	17	3,933	171
October	4,439	193	230	10	4,669	203
November	3,772	164	483	21	4,255	185
December	2,277	99	207	9	2,484	108
Sub-Total	49,864	2,168	5,888	256	55,752	2,424
Total	86,756	3,772	9,867	429	96,623	4,201

TEACHING NOTE EXHIBIT 2
Least-Squares Regression Analysis

Least-Squares Regression Analysis									
Computation of Regression Estimates					Computation of R ²				
Year	Month	Utility Cost for the Month	Kits Sold			Predicted Cost Based on Regression Line			
		Y	X	X ²	XY	Y ¹	(Y-Y ¹) ²	(Y-Y ¹) ²	
2011	April	496.70	249	62,001	123,678.30	427.965	4,724.555	6827.402	
	May	472.94	229	52,441	108,303.26	422.289	2,565.564	3465.453	
	June	494.19	189	35,721	93,401.91	410.937	6,931.131	6418.909	
	July	447.27	143	20,449	63,959.61	397.882	2,439.197	1102.114	
	August	379.90	251	63,001	95,354.90	428.532	2,365.091	1167.719	
	September	386.08	190	36,100	73,355.20	411.220	632.039	783.547	
	October	366.46	255	65,025	93,447.30	429.667	3,995.176	2266.893	
	November	358.34	157	24,649	56,259.38	401.855	1,893.553	3106.045	
	December	461.88	114	12,996	52,654.32	389.652	5,216.947	2285.614	
	2012	January	613.06	226	51,076	138,551.56	421.437	36,719.298	39596.262
		February	498.66	201	40,401	100,230.66	414.342	7,109.493	7155.146
		March	498.41	263	69,169	131,081.83	431.938	4,418.552	7112.914
April		457.22	229	52,441	104,703.38	422.289	1,220.203	1861.758	
May		375.27	254	64,516	95,318.58	429.384	2,928.282	1505.588	
June		396.90	218	47,524	86,524.20	419.167	495.810	294.874	
July		360.58	159	25,281	57,332.22	402.423	1,750.801	2861.384	
August		342.05	207	42,849	70,804.35	416.045	5,475.259	5187.155	
September		293.38	171	29,241	50,167.98	405.828	12,644.594	14566.536	
October		239.30	203	41,209	48,577.90	414.910	30,838.799	30545.219	
November		335.87	185	34,225	62,135.95	409.801	5,465.850	6115.538	
December		421.05	108	11,664	45,473.40	387.949	1,095.692	48.694	
		8695.51	4,201	881,979	1,751,316.19	8,695.510	140,925.887	144274.763	
	Mean	414.072							
	Intercept	357.298							
	Slope	0.2838							
	R ₂	0.02321							

ASB 2013
Antigonish, Nova Scotia

Dr. Mark Fuller
Dr. Stefan Litz
St. Francis Xavier University

**TIES THAT BIND:
AN INSTITUTIONAL ISOMORPHIC COMPARISON OF MANDATORY
CSR REPORTING STANDARDS IN THE CANADIAN FINANCIAL SERVICES
INDUSTRY**

ABSTRACT

This empirical paper represents a novel study of a mandatory CSR reporting regime. Five Canadian financial institutions are compared in terms of their references to three voluntary reporting standards in their annual and CSR reports. Institutional isomorphism is used as our frame of reference to examine how CSR reporting varies across these firms. Implications arising from this study affect the future development of CSR standards, the practices of CSR managers, and advance theory for CSR researchers.

INTRODUCTION

Mandatory corporate social responsibility (CSR) reporting is a relatively new phenomenon in Canada. It only became a requirement for Canada's biggest banks in 2001, following the passing of Bill C-8, *An Act to establish the Financial Consumer Agency of Canada* (Canada, 2001), which among other requirements made the issuing of public accountability statements mandatory for financial institutions with equity exceeding \$1 billion. Prior to this Act, such reporting was voluntary, and it remains so for firms outside of the financial services sector, and even for financial institutions with smaller equity holdings. With voluntary reporting, firms that can afford the costs associated with collecting and disseminating CSR information may choose to differentiate themselves from rival firms that opt not to follow suit. From a research perspective, this can result in methodological challenges in terms of the sample size and quality. Whereas with a mandatory reporting regime, the differences among firms when every organization has a compulsory obligation to report enable better comparisons into both intra-industry firm behavior and the CSR standards themselves.

Our research objective is to examine how CSR standards with differing origins and characteristics are incorporated into the CSR reporting activities of large firms. We chose as the sample for our study the Canadian financial services sector, and undertook an in-depth 12-year study of the five largest institutions' CSR reporting activities. These firms operate under a mandatory CSR reporting regime, in contrast to other firms in Canada, and to financial institutions in other major countries such as Germany, the United Kingdom and the United States. This facilitates intra-industry comparisons that would otherwise not necessarily be possible in industries with a voluntary CSR reporting regime in which some firms choose not to report. Facing the same regulatory requirements to report, but with great liberty with what to

report under Statutory Regulation 2002-133 (Canada, 2002), financial institutions in Canada began mandatory reporting in 2001/02.

We have collected both the annual reports and the public accountability statements for these five institutions for the period 2001 to 2012 and using content analysis, identified the three most frequently cited CSR reporting standards: the Equator Principles, the United Nations Global Compact, and the Global Reporting Initiative. For these three reporting regimes, we next examined the nature and the frequency with which each reporting standard was measured. We qualitatively analyzed each mention of the reporting standard and in order to assess the degree to which these standards were employed in firms' CSR and annual reports. We begin with a discussion and comparison of these three standards, followed by a review of the literature. The research methods by which this study was undertaken are next detailed, and this is accompanied with the findings of our qualitative analysis. We conclude by articulating the implications of this study for practitioners and academics.

Literature Review

In our review of the extant literature, we wish to address two issues. The first issue is why an understanding of CSR reporting behavior is a worthwhile research endeavor. The second are the incremental benefits of our particular study; how it builds upon the scholarly work that came previously.

Importance of this Research Topic

The value in a study of CSR reporting behavior is premised upon the value-added contribution of existing research. This research field of inquiry has concluded that CSR activities provide a range of benefits to firms which justify the reporting effort. Various studies (Adams and Ambika, 2005; Khan, Halabi, and Martin, 2009) suggest these benefits include improved employee retention and recruitment (Simms, 2002); enhanced business decision-making and cost reduction (Adams, 2002; King, 2002; Simms, 2002); strengthen corporate image and stakeholder relationships (Marx, 1992/93; Pike, 2000; Adams, 2002; Bernhut, 2002; Joyner and Payne, 2002; King, 2002); and improved financial performance (Margolis and Walsh, 2003). Additional benefits include improved transparency and perceptions to be less involved in corporate malfeasance (Tapscott and Ticoll, 2003). As a result, the last decade has seen the development of a number of CSR reporting initiatives, frameworks and standards which may be adopted by corporations for their CSR reporting. For this study, we reviewed 125 CSR and annual reports for five Canadian financial institutions that contained references to a plethora of CSR standards. In a process detailed in the *Research Methods* section below, the three most referenced CSR standards among leading Canadian financial institutions were identified. These were the Equator Principles, the United Nations Global Compact, and the Global Reporting Initiative. Studying firms' references to these CSR reporting standards, of which these three are examples, helps practitioners improve their CSR reporting activities and researchers to:

- Understand how differences in CSR reporting standards are associated with differences in CSR reporting activity; and

- Discover how firm behavior in terms of CSR reporting activity changes over time.

Incremental Benefits of this Study

The second aspect of our literature review that we wish to address is what incremental benefit can be gained from this particular study of the topic. We identified 18 peer-reviewed scholarly publications of potential relevance to our study that made reference to the Equator Principles, the UN Global Compact and the Global Reporting Initiative. Of these articles, none provided an empirical comparison of these CSR standards in terms of firms' corporate reporting activities. Three were relevant in other contexts. For example, Jamali (2010, 617-618) noted the existence of a research gap in this regard, writing that "while there has been a proliferation of academic writings, examining IAS [International Accountability Standards] from the perspective of their scope, purpose, and overall utility, the practical engagement of practitioners and MNCs with these standards has not been systematically carried out, despite the fact that they constitute the audience for whom these standards are most intended." Pérez and Rodríguez del Bosque (2012) highlighted the importance of the financial services sector as a medium for CSR research, finding significant CSR investments by the financial services sector in their study of six Spanish financial institutions and how CSR activities shape corporate personality and how firms' organizational behavior influences their CSR reputation. Their findings further support our premise of the validity of the financial services sector as an appropriate venue for CSR-related research. Cuesta-González, Muñoz-Torres, and Fernández-Izquierdo (2006, 290-291) noted the central role of banking institutions to the economy and commented that "new management strategies with sustainability criteria that companies currently encounter have forced banks to analyze their role in society and their contribution towards obtaining more sustainable development." They classified research on corporate social responsibility in the banking sector into three categories: theoretical approaches tying together economics and business ethics; CSR behavior and the relationship to social and financial performance; and the design and implementation of reporting systems to assess social performance, of which this article falls into the third category.

We next examined the literature pertaining to research methods and contexts regarding CSR reporting and analysis in the financial services sector. Research methods have typically emphasized content analysis, including the empirical review of annual reports, (Khan, Halabi and Martin, 2009; Khan, 2010; Bayoud, Kavanagh and Slaughter, 2012; Mondal and Ghosh, 2012), CSR reports, or both (Idowu and Towler, 2004). Among the many geographic contexts examined, studies have been undertaken in Australia (Cuganesan and Khan, 2008); Bangladesh (Khan, Halabi and Martin, 2009; Khan, 2010), India (Mondal and Ghosh, 2012), Libya (Bayoud, Kavanagh and Slaughter, 2012) and the United Kingdom (Idowu and Towler, 2004). In terms of the subject context, Nikolaeva and Becho (2010) claimed to have the first study to examine the voluntary adoption by firms of Global Reporting Initiative CSR standard. They incorrectly state that "at least up to the end of 2009, there are no regulatory bodies in any country that require companies either to undertake or report on CSR activities" (Nikolaeva and Bicho, 2010, 138) when in fact government imposed mandatory reporting in the Canadian financial services sector has existed since 2001 (Canada, 2001, 2002). What our literature review has revealed is that a comparative empirical examination of CSR reporting standards in both annual report and CSR reports has not been undertaken, particularly with respect to the interesting case of Canada with its mandatory CSR reporting regime for large financial services firms. By addressing this

knowledge gap from an institutional perspective, this paper provides an incremental scholarly contribution to the CSR field.

Research Methods

An initial caveat for the research was that the authors intended to study an industry in which CSR reporting was mandatory. The necessity of this requirement was to foster comparisons among firms within a single industry; to enhance the quality of the analysis; to reduce sample bias; and to provide a starting point for comparisons to other industries and countries where CSR reporting is of a voluntary nature and where the data is less complete. The banking sector in Canada was chosen as this segment of the industry is federally regulated and one of the regulations mandates CSR reporting in the form of public accountability statements (Canada, 2001, 2002). This mandatory CSR reporting environment provided an excellent data source for our study.

Data Sample

We began our study by identifying the top five Canadian financial institutions based upon total assets, all of which surpass the mandatory reporting threshold established by the federal government (Canada, 2001). We have used the following monikers to make their identities anonymous: Bank Alpha; Bank Bravo; Bank Charlie; Bank Delta and Bank Echo. We collected the annual reports for each firm from 2000 to 2012. We next collected the corporate social responsibility reports for each firm's Canadian operations, some of which began in 2001 and others in 2002, through 2012. In most cases, firms in a given year published their annual report and their CSR report as separate documents. There were two exceptions when:

- A financial institution produced a CSR report that was separate and generally broader in scope than the public accountability statement which they also published; or
- The firm integrated their CSR report and annual report into a singular document.

This produced a list of 125 documents spanning the time period of 2000 to 2012.

Identification of CSR Standards

We next identified every CSR standard mentioned during this time period in any of the reports. Once a comprehensive list of standards was developed, we then counted the number of references to each standard, and the number of documents in which each standard was measured. We then calculated the mean number of references per document for each standard. The three most frequently mentioned standards, according to this measure, were the Equator Principles, followed by the United Nations Global Compact and the Global Reporting Initiative; these three CSR standards became the basis for this analysis. This data collection process intentionally excluded any standalone reports on any of the standards, and corporate websites, in order to enhance inter-firm comparisons within the industry.

Comparison of CSR Standards

This selection process resulted in the analysis of three very different CSR standards. The Equator Principles (2006) are relevant to those firms engaged in project financing, which is relevant in

the financial services sector but not every industry sector. The UN Global Compact (United Nations 1999, UNGC 2011) is multi-sectoral by nature and is designed to promote ethical business principles. The Global Reporting Initiative is also multi-sectoral, but is designed around specific reporting requirements for each of the dimensions it examines; in comparison to the Global Compact, the GRI lends itself to being operationalized by firms to a much greater degree (GRI, 2011a). It also features a supplement applicable to the financial services sector (GRI, 2011b). We summarize the key components of each CSR standard in Table 1.

Insert Table 1 here

Analytical Methods and Validity

A case analysis-based approach to the research methods was employed. Each of the five firms was treated as a single case, with the subsequent analysis consisting of within-case and cross-case analyses (Eisenhardt, 1989; Miles and Huberman, 1994) using each report as the basic unit of analysis. Within-case analysis involves the documentation and review of individual case descriptions as distinct entities. Cross-case analysis invokes a comparative approach across multiple cases to identify areas of commonality and difference. CSR and annual reports for each of the five firms were individually and collectively analyzed using a four-item ordinal scale. A coding key was developed a priori analysis in order to promote the consistent assessment of each report that was reviewed, and was revised as needed during the analysis, which was then repeated, to achieve accuracy and reliability in the coding (Weber, 1990), and to promote intersubjective verifiability.

Ordinal Rating:	Description:
A	Fulsome discussion or analysis of the CSR standard for the particular reporting year. Examples include the applied use of a GRI Index, Equator Principles-specified reporting template (2011), or a detailed discussion of the UN Global Compact principles and their application by the financial institution.
B	CSR standard discussed in the context of strategic plans or objectives.
C	CSR standard discussed in the context of the firm's operations, policies of activities.
D	Nominal mention consisting of little descriptive information. A passing reference.

Firms were assigned anonymous labels: Bank Alpha, Bank Bravo, Bank Charlie, Bank Delta and Bank Echo. CSR reports were analyzed first, in sequential order, followed by annual reports, as the former tend to have broader discussions of the CSR standards. This enabled the authors to better assess the extent of the information available for inclusion in the annual report, which typically includes a synthesis of key points from the CSR report. Each report was analyzed in terms of the three CSR standards: the Equator Principles, the UN Global Compact,

and the Global Reporting Initiative. All mentions of these standards were reviewed in terms of the context and significance of their use. Since the references to each of the three CSR standards represented a small portion of the overall content of each CSR and annual report, and we were interested in comparing changes in CSR reporting in aggregate, with the report being the basic unit of analysis, we chose to employ whole text coding (Weber, 1990) for each report. To advance inter rater reliability, we opted to escalate the ordinal ranking assigned to a report to the highest rating for which any element of the report, individually or collectively, was able to achieve. This rule resulted in the benefit of the doubt being given to each report: it favoured the assignment of a higher rating, rather than a lower rating, in order to reduce the potential for rating inconsistencies arising from the assignment of ratings.

References to the CSR standards, once classified according to the scale above, were logged into a table to facilitate comparisons among CSR standards for each firm, across firms, and across standards. We quantitatively assessed the results of our qualitative ratings of each CSR and annual report. We created frequency distributions for the quality ratings of each report, by CSR standard and by each financial institution. We also examined the frequency of both total references to each CSR standard, as well as the frequency of non-mentions, in order to evaluate what was written as well as what was not written in the reports. We then segmented our findings in terms of both the CSR standard – to give insight on each reporting mechanism – and by financial institution – to give insight into intra-industry dynamics. Results of this analysis are reported below.

Findings

We report our findings in two sections. In the first section, we examine CSR reporting in terms of the three standards examined. In the second section, we explore the topic from the perspective of firms in the financial services sector in Canada.

Findings by CSR Standard

Of the three CSR standards, the Global Reporting Initiative (GRI) originated first in 1997. However, the first set of standards was not released until three years later (GRI, 2011c). That same year, the United Nations Global Compact was released, followed by the Equator Principles in 2003. For each reporting year (2000 to 2012 for the GRI and the UN Global Compact, 2003 to 2012 for the Equator Principles), we examined whether and which institutions made reference to the standard, and classified the extent to which the standard was mentioned in each of their CSR and annual reports.

Among our initial findings was that there was a delay between the introduction of a standard and the first references being made by a Canadian financial institution. For example, the Global Reporting Initiative was introduced in 2000 but it was not until two years later, in 2002, when one bank first referenced it in either an annual or a CSR report. It was 2006 – a full six years from launch, that the UN Global Compact was referenced. By comparison, the Equator Principles was first mentioned in the year of its launch, by three financial institutions. This may be a function of the perceived relevance of the standard to a firm, as the Equator Principles are industry-specific to financial service providers, whereas the UN Global Compact is a pan-industry, principle-driven standard. As slow as some institutions were in even acknowledging

this particular global CSR standard, it is noteworthy to mention that one institution (name withheld) has not included any reference to the United Nations Global Compact in any of its annual or CSR reports from 2000 to 2012. From an institutional perspective, this is counterintuitive: while strategic responses to institutional processes include resistance toward pressure to conform (Oliver, 1991; Jamali, 2010) over the long term, isomorphic forces (DiMaggio and Powell, 1983) would be expected to eventually result in increasingly similar reporting practices.

Our second finding was in relation to the rate of diffusion of the various CSR standards. For the Global Reporting Initiative, it took a mere two years from the first bank to the last bank to make at least a passing reference to the standard (2002 to 2004). The Equator Principles took twice as long: 2003 to 2007. Worse still was the UN Global Compact: While one institution voiced support for these principles in 2006, and four of the institutions acknowledged it by 2010, one of the financial service providers still has yet to acknowledge the standard in any of their annual or CSR reports. This indicates a considerable time lag between the introduction of a CSR standard; its first reference or inclusion in a firm's CSR reporting; and widespread inclusion across all firms in an industry. This finding adds to the theoretical uncertainty of CSR reporting, adding issues of temporal difference in the effects of institutional isomorphism (DiMaggio and Powell, 1983; Oliver, 1991; Jamali, 2010) on firms' CSR reporting behavior.

We next drilled down in the data for further insights about the frequency and quality of references to each of the three CSR standards. Each financial institution had 26 opportunities to mention either the GRI or the UN Global Compact, and 20 opportunities to mention the Equator Principles during the time period studied. Collectively, this results in 130 possible inclusions of the GRI and UN Global Compact standards, and 100 inclusions of the Equator Principles, for the five banks which we examined. Our findings reveal that in terms of total references to each standard, the Equator Principles was referenced in 78% of the CSR and annual reports; the GRI in 48% of the reports; and the Global Compact in just 8% of the available opportunities. The Equator Principles also had the lowest frequency of passing references: just 15% of all mentions were assessed as of marginal value. However, the GRI had the highest frequency of top quality references – 40% of all references were deemed to represent a fulsome analysis or discussion of the standard. The CSR standard with the lowest inclusion was the UN Global Compact: there were only 11 reports out of 130 that made reference to this standard; none of the references were of a top tier nature; and 73% of the references were of marginal value. These results indicate that while the Equator Principles was the most referenced standard, and the UN Global Compact the least referenced, with the GRI representing the CSR standard that was best incorporated into financial institutions' CSR and annual reporting in terms of the quality of reporting. These variations in the frequency and quality to which each CSR standard is referenced adds to the theoretical ambiguity of the role that isomorphic forces play in CSR standard implementation.

Findings by Financial Institution

In examining the results for each financial institution, we totaled each of the number of A, B, C, and D-rated mentions for the three CSR standards. This represents 26 reporting opportunities to mention either of the GRI or UN Global Compact standards, and 20 opportunities to mention the Equator Principles. We found an equally poor level of reporting performance for each firm during the time period studied. Bank Delta failed to mention any of the three CSR standards 56% of the time in 72 opportunities. Bank Alpha failed to mention the

standards 57% of the time; Bank Bravo and Bank Charlie 58% of the time and Bank Echo 60% in 72 opportunities. These results are all attributable to infrequent reporting in the early years of this study and the frequent non-reporting of references to the United Nations Global Compact. This finding is evidence of isomorphic forces at work, as the number of references to CSR standards did increase for all financial institutions after the initial time period where poor reporting frequencies were noted.

We also examined the results by financial institution in terms of the number of actual references to each standard. In this way, we assess each firm in terms of what they did report, not what they did not report, in the context of the three CSR standards. The five firms had a remarkably similar number of reports in which the CSR standards were mentioned; these ranged from a low of 29 reports for Bank Echo to a high of 32 reports by Bank Delta. Firms in the Canadian financial services sector have not, to date, differentiated themselves in terms of the frequency of CSR reporting. This is further evidence of institutionalized isomorphism: we would not expect, over the long term, for there to be substantial differences in the total number of published reports, as the inconsistent reporting activity becomes marginalized with the passage of time.

The pattern of citation across institutions and over time was also of interest. We were interested to see which standards were most consistently mentioned from CSR to annual reports, year in, year out. For Bank Alpha, the GRI was mentioned consistently in both types of reports from 2007 to 2012, the last year for which data is available. Bank Echo also more consistently referenced the GRI than the other standards, mentioning it unfailingly from 2007 to 2012 in both report types. For their three rivals, each mentioned the Equator Principles more consistently, with Bank Bravo having the longest streak, from 2004 to 2012 inclusive. This suggests a differentiated approach to CSR reporting, and possibly to CSR behavior, with some banks emphasizing a broad spectrum approach to the topic (those that favour the GRI) and those that take a more project finance perspective (those that focus on the Equator Principles). While this is counterintuitive from an isomorphic perspective, it is consistent with the findings of Pérez and Rodríguez del Bosque (2012) who also found the firms employ CSR practices differently in order to shape and influence external perceptions of their corporate identity.

One source of differentiation from firm to firm was the quality by which the CSR standards were incorporated into the CSR and annual reports. In each year that Bank Charlie issued either an annual or a corporate social responsibility report, 47% of the time that report contained a thorough incorporation of one of the three CSR standards we studied. Bank Alpha scored lowest in terms of top quality references to CSR standards, a mere 13% of all references by that institution; the industry average was 30%. We also looked at the frequency of passing references to CSR standards. In this case, Bank Bravo had the lowest frequency of marginal references to each of the CSR standards, at 23%. Bank Echo had the highest percentage of passing references, at 55%, which was well above the industry average of 35%. This reflects a great disparity in the quality of the CSR reporting from institution to institution. However, this measure is for the entirety of the timeline studied, 2000 to 2012. To more fairly gauge the theoretical significance of this finding, we need to see whether quality improvement is or is not happening within the time horizon studied.

Thus, we wished to examine whether the CSR reporting performance of firms improved over time. If quality improvement in CSR reporting is occurring, this would be evidence of

organizational learning and development (Argyris and Schoen 1978; Argyris 1982), which would be consistent with a rise in homogeneous CSR reporting activity within the industry as a result of isomorphic forces. To explore this issue, we chronologically segmented our quantitative results into three time periods: the introductory period of 2000 to 2004, which includes the first mention of one or more of the standards by each financial institution; the growth period of 2005 to 2009 in which CSR reporting became more frequent; and the contemporary period of 2010 to 2012 which incorporates the most recently published data. Our intent was to see whether the frequency of A-quality top tier mentions improved over time. This would imply a movement toward more substantive CSR reporting over time. To adjust to the different lengths of time in each period (5 years, 5 years and 3 years respectively), we computed results on a percentage basis. In terms of A-rated references to the CSR standards, four of five firms continuously improved the percentage of top tier references in each time period; only Bank Delta saw a decline in the most recent time period of 2010 to 2012. There was also a substantial variation in the frequency of the top tier references, with 17% of Bank Alpha reports in the final time period were of a superior caliber, whereas 75% of Bank Charlie references to the same standards were exemplary. This suggests that for the majority of banks, the frequency of their top-tier CSR reporting was becoming more frequent. Consistent with our findings, we would theoretically expect that as CSR knowledge and the application of best reporting practices becomes more diffusely spread within the industry, the incidence of A-rated references would increase and the frequency of D-rated references would decline.

As part of our assessment of organizational learning and development, we also investigated whether the frequency of D-rated passing references declined over time. This would suggest a movement away from less substantive mentions of the CSR standards over time. Of the five institutions, only Bank Alpha exemplified the expected pattern, with a declining frequency of passing references in each of the three time periods. Bank Bravo had a moderate frequency of passing references initially (40%), declined to a healthy 13% in the interim period, before rebounding to 33% of all references being passing references in the contemporary period. Bank Charlie began with no passing references, surged to 32% in the growth period for CSR reporting, and then reversed course slightly and finished at 25% of contemporary references to the CSR standards being passing ones. Bank Delta and Bank Echo had initial declines in D-rated references before experiencing slightly increased D-rated references in the contemporary period. One can speculate that these mixed results may reflect differences in perception as to whether the CSR standards are sufficiently embedded in stakeholder thought so as to not require as substantive a discussion as was needed previously, or whether there has been an organizational shift to using buzzwords over substantive discussions of CSR-related issues.

Additionally, we examined the frequency of non-references to the CSR standards. A non-reference occurs when a financial institution issues a CSR or an annual report and does not mention each of the three most prominent CSR standards despite having had three opportunities to do so with each report. Unlike examining the frequency of D-rated references, this calculation does not factor quality into the analysis, only the presence of *any* mention of the CSR standard. Thus we would expect the frequency of non-references to decline in each time period as the CSR standards become more embedded within the organization. By this measure, three of five institutions exemplified the expected pattern: Bank Alpha, Bank Delta and Bank Echo. Both Bank Bravo and Bank Charlie showed initial declines in non-response from the introductory time period to the growth time period, countered by increases (slight for Bank Bravo, more substantial

for Bank Charlie) in the contemporary time period. As a result of these three assessments of organizational learning and development, Bank Alpha demonstrated the most consistent improvement: their A-rated references continuously improved and the incidence of both D-rated reference and non-references declined in tandem. All other firms' results were mixed. This is not a pattern we expected from a theoretical perspective; greater consistency in the decline of non-references was anticipated.

To conclude our examination of organizational learning and development, we offer a final contribution to this topic, by classifying firms according to the relative frequency with which they engaged in A-rated and D-rated CSR reporting of the three preeminent reporting standards. To assess the frequency with which CSR standards were mentioned in a firm's CSR reports, we calculated the total number of A-rated references to the three CSR standards for the introductory CSR reporting period, 2000 to 2004. We then found the median number of A-rated mentions CSR standards occurring in all firms' CSR and annual reports for this time period. If the firm's total number of mentions was greater than or equal to the median number of mentions for the industry, we classified this in the "high" category, which represented firms that more actively engaged in high quality inclusions of the CSR standards than their peers in the sample; otherwise, the firm was classified in the "low" category. We replicated this approach with D-rated or passing reference to the three CSR standards in each firm's CSR and annual reports from 2000 to 2004, and similarly classified them as "high" or "low" based on the median of the industry. As a result, we were able to classify each of the five financial institutions in terms of their reporting frequency of the CSR standards, as indicated in the table below:

Insert Table 2 here

The table above parses industry members by the quality of their CSR reporting during the introductory reporting period. Bank Charlie is in the optimal position, with a high frequency of in-depth references to the CSR standards, and a low frequency of CSR reporting involving passing references to these standards. No firms are in the second most optimal position, the bottom right quadrant, in which a firm produces an above average level of both A-rated and D-rated references to the standards. Bank Bravo and Bank Delta, while not embedding the three CSR standards into their reporting as effectively as Bank Charlie does, at least curtail the frequency of fluffy, non-substantive references to these standards. Unfortunately, both Bank Alpha's and Bank Echo's initial CSR reporting were found to be lacking in depth of analysis of the CSR standards and a proclivity to engage in the use of passing references. We repeated this analysis for the most recent reporting period, 2010 to 2012, to see whether there was evidence of improvement.

Insert Table 3 here

From Table 3, we can see that Bank Charlie maintains its position in referencing the three CSR standards to a high degree, while minimizing the use of passing references when compared to their peers. Bank Bravo has improved their reporting and joins Bank Charlie in this category.

Bank Delta, while not embedding as substantive references to the CSR standards in their reporting as Bank Bravo and Bank Charlie, still manages to refrain from significant use of passing references. The firm is joined by their peer, Bank Alpha, in this category. Unfortunately, Bank Echo continues to engage in low-quality referencing of these standards relative to their peers, and in the use of passing references to these important CSR standards. This is not to suggest that Bank Echo is not improving, but that the pace of their improvement is not on par with the pace of change in rival institutions. This suggests that isomorphic forces are experienced differently, and responded to differently, by firms. This finding is consistent with the literature on institutional isomorphism (Oliver, 1991; Jamali, 2010).

Implications

As a result of the findings from this exploratory study suggest, there are implications for CSR advocates, practitioners, and researchers we wish to discuss.

Implications for CSR Advocates

For CSR advocates, an important observation relates to the successful launch of a social reporting standard. In the continually evolving world of social reporting, there are a plethora of multi-sectoral reporting standards that have emerged, such as the Global Reporting Initiative and the United Nations Global Compact. While the potential for new multi-sectoral standards may be limited, there is also the potential for sector-specific standards, such as in the case of the financial services sector with the Equator Principles, to emerge in the future. Thus it is not unreasonable to expect new CSR standards to develop in the years ahead in sectors beyond the financial services realm.

With this in mind, we have noted that CSR standards appear to be more readily adopted by a particular firm with there are business endorsements at launch in the same sector, than standards launched either without business endorsements or endorsed by firms from a different sector than that of the prospective participant. This observation arises from the qualitative evidence which suggested a relationship exists between the breadth of business endorsements for a CSR standard at launch and the length of time before major financial institutions chose to participate in such standards. Ten international financial institutions signed on at the launch of the Equator Principles in 2000 (Equator Principles, 2013). Of these ten firms, two Canadian firms – Bank Charlie and Bank Bravo – signed on within the first few months of launch. With both Canadian firms making reference to the Equator Principles in their CSR reports that first year, Bank Alpha and Bank Delta followed suit two years later, with Bank Echo joining in 2007. In comparison, the UN Global Compact was endorsed at its launch by 35 multinationals, including six non-Canadian financial services firms (UN Global Compact, 2012). Since then, only Bank Delta in 2010 has formally signed on as a participant and to date Bank Bravo has yet to even acknowledge it in either their CSR or their annual reports. Similarly, the Global Reporting Initiative involved no financial institutions at its founding in 2000, and it was not until two years later before any of the financial institutions examined in this study made reference to it in their annual or CSR reports.

For CSR standards to avoid delays in making a positive impact for social change, we posit that negotiating for corporate partners prior to the time of launch, possibly key players or

flagship companies (Rugman and d’Cruz, 2000) in a variety of industry sectors is a key success factor. This is consistent with the literature, which suggests that early adopters of a CSR reporting standard or framework, particularly those at launch, implement the principles of a given CSR standard to a greater degree than latter adopters, while at the same time being a source of influence and support for the framework (Arevalo, 2008). We extend this proposition to suggest that organizations which have launched CSR standards actively work on business partnership engagement to maintain and grow their roster of firms which are employing their particular CSR standard, as competition for the endorsement of standards by flagship firms will likely increase.

Implications for CSR Practitioners

Our study noted the following findings on a firm-level basis:

- Firms have become consistent in their referencing of the Equator Principles and the Global Reporting Initiative by the end of 2013;
- Firms have not differentiated the frequency of their CSR reporting; each issues a CSR report annually;
- Differentiation in the content of CSR reporting has occurred, with some firms (Bank Alpha and Bank Echo) emphasizing broad-based measure of CSR while others (Bank Bravo, Bank Charlie and Bank Delta) choosing a more narrowly focused perspective;
- The quality of the CSR reporting varies among firms both in terms of the frequency of high-quality reporting and the frequency of low-quality reporting, which are two distinct variables;
- Organizational learning and development in terms of CSR reporting is occurring. In absolute terms, four out of five firms (excluding Bank Delta) saw a continuous improvement in the frequency of high quality A-rated references to the three CSR standards. Only one firm, Bank Alpha, saw a continuous decline in the frequency of low quality D-rated references. Three firms – Bank Alpha, Bank Delta, and Bank Echo – consistently reduced the frequency of non-references to the CSR standards. On a relative basis, Bank Charlie has been a top tier performer in terms of their relative CSR reporting, and has been joined by Bank Bravo that has experienced improvement. Bank Alpha has improved to a mid-range level, joining Bank Delta whose position relative to the industry is unchanged. Bank Echo remains the poorest performer in terms of CSR reporting quality compared to industry rivals.

With institutions having established a yearly reporting frequency for their CSR reports in response to the public accountability report obligations, our advice to CSR practitioners is to continuously improve the quality of their CSR reporting in these two reporting venues. Performance improvements may arise from improving not only the frequency of high-quality, detailed references to the CSR standards in both their report venues, but reducing the frequency of passing and non-references to the standards. Additionally, firms may wish to differentiate themselves in terms of the frequency of CSR reporting. While each firm has supplemental websites and/or additional CSR-related materials, it is the CSR reports that are read by the CSR community, as it is the annual reports read by the investment community; firms should therefore

be careful not to fragment their CSR efforts by reporting on their activities in different media without synthesizing it in a single comprehensive repository, which is what the CSR report is best positioned to do. Lastly, while each firm studied showed indications of improvements over time, the pace of that improvement did vary among financial institutions. For firms such as Bank Alpha, Bank Delta and especially Bank Echo, there is an opportunity to improve their performance to Bank Bravo and Bank Charlie performance levels. For Bank Bravo and Bank Charlie, to remain ahead of their rivals will be a challenge since these two firms lack a superior quality benchmark for which they may aspire; maintaining their lead will be difficult when the performance metrics of future CSR success remain ill-defined.

Implications for CSR Researchers

For CSR researchers, much effort has been spent on analyzing the content of CSR reports and related activities (Chapple and Moon, 2005; Reynolds and Yuthas, 2008; Vurro and Perrini, 2011). While we concur that such analysis is important, and makes a valuable contribution to the social reporting literature, we suggest additional effort is merited in how frequently, and consistently, is that content reported. From our analysis, it would appear that there are many inconsistencies with regard to the nature and frequency of the CSR reporting by firms. Some firms are quick to join new CSR reporting standards, others slower and for some firms, they resist the actions of their rivals within the industry, suggesting a constraint on the isomorphic effects within the financial services sector as it pertains to corporate social responsibility reporting. Even when a firm undertakes to report according to a particular CSR standard, that reporting may be limited to their CSR reports and may or may not extend to inclusion in their annual reports; this phenomenon brings into question the institutional embeddedness of CSR activities. Further, once reporting has begun, it does not imply it will continue uninterrupted, as a number of firms displayed inconsistent reporting behavior regardless of the reporting standards which they employed; this is suggestive of changing perspectives with regard to emerging norms of CSR reporting activities. As a result, questions arise from an institutional theory perspective as to the impact CSR reporting has had on firm behavior. Further research is merited in this regard: to better assess the extent to which CSR values and principles are truly embedded within an organization, as opposed to merely articulated for public relations purposes.

From an academic perspective, this paper contributes to the discussion about institutional isomorphism and firm behavior. While the literature has established that firms become more similar over time (DiMaggio and Powell, 1983) while experiencing different firm responses to isomorphic pressures (Oliver, 1991; Jamali, 2010), our findings suggest either of two possibilities: norms of CSR reporting activities are still developing and thus rival firms are not perceiving their own outlying behavior as being counter to established practice, or, firms are intentionally resisting isomorphic pressures in their CSR reporting activities because of perceived benefits from doing so. Further research is necessary to resolve this issue, and with the passage of time, an increase in the availability of annual and CSR reports should provide much needed clarity.

Discussion

The purpose of this paper was to improve our understanding of how CSR standards are implemented in the annual and CSR reporting activities of large firms. We chose the financial

services sector in Canada because federally regulated banks with equity over a particular threshold must engage in CSR reporting by law; a condition that is evidently quite rare in international circles (Nikolaeva and Bicho, 2010). In our examination of five Canadian banks, we found both similarities and differences in terms of (a) the implementation of different CSR reporting standards; and (b) CSR reporting activity by firms. From the perspective of developing and embedding voluntary CSR reporting standards, we posited the following:

- New CSR reporting standards are likely to be developed in the future;
- Those reporting standards with business endorsements at launch are more readily adopted by firms than those that have fewer or no endorsements; and
- Competition for endorsement by flagship firms will necessitate greater business partnership engagement.

Consequently, we anticipate that the existing fragmented nature of CSR standards will eventually give way to an environment where a few CSR standards, such as the Equator Principles and the Global Reporting Initiative, are dominant, and other standards become marginalized or disappear altogether.

From the viewpoint of financial institutions, we found that firms are, or have become, more alike in terms of their use of two of the three CSR reporting standards and the frequency with which they issue their annual and CSR reports. We also discovered that firms are unlike in the nature of their CSR activities; the quality of their CSR reporting; and the pace of their organizational learning and development. Given the newness of this organizational activity, we expect firm behavior to continue to evolve as norms become established and proliferate in regards to CSR reporting. However, much evidence exists of the rising impact of isomorphic pressures in the field of CSR reporting. For firms wishing to compete against their better positioned rivals, opportunities for effectively using social competition instead of, or in addition to, economic competition are shrinking with each passing year.

We finally wish to acknowledge three particular limitations of this study. The first is in relation to the data sample. For the purposes of consistency, we only evaluated the CSR/public accountability reports and the annual reports for five institutions that were publicly available. While this did entail a sample of 125 reports, some financial institutions have referenced one or more of the CSR standards in topic-specific reports and websites. Internal, non-public reports may also exist. As we could not determine with certainty the population of these irregular reporting activities, we chose the more conservative approach of analyzing just those report types published by all firms and which are in the public domain. A second limitation relates to the ordinal scale for our assessment of each CSR or annual report. While we have applied the report assessment scale consistently across years and firms, scales that involve qualitative assessments are subjective, and therefore open to issues of inter rater reliability. A change in scale or rater could conceivably impact the outcome of the study. Last is that the study is confined to Canadian firms reporting in a mandatory CSR reporting regime. While such a scenario facilitated data access for this study, the results may not be generalizable to non-Canadian geographies, or for industry sectors with voluntary reporting regimes both inside and outside of Canada.

Table 1: Comparison of the Three CSR Reporting Standards

	Equator Principles	United Nations Global Compact	Global Reporting Initiative
Introduction	2003	2000	2000
Scope	International	International	International
Revisions	Yes, EP-1 2003; EP-2 2006, EP-3 2012.	None.	Yes, GRI-1 2000, GRI-2 in 2002 and GRI-G3 issued in 2006. GRI-G4 published in 2013.
Dimensions	<ol style="list-style-type: none"> 1. Social and Environmental Risks 2. Labor and Working Conditions 3. Resource Efficiency and Pollution 4. Community Health and Safety 5. Land acquisition and involuntary resettlement 6. Conservation of Biodiversity 7. Indigenous People 8. Local Cultural Heritage 	<ol style="list-style-type: none"> 1 Human Rights 2 Labor Conditions 3 Environmental Impact 4 Corruption 	<ol style="list-style-type: none"> 1 Economic Performance 2 Environmental Performance 3 Social Performance <ol style="list-style-type: none"> 3.1 Labor Practices 3.2 Human Rights 3.3 Society 3.4 Product Responsibility
Independent Third Party Assurance Incentives or Requirements	No	No	If the GRI report has been assured by a independent company the GRI permits the company to add a (+) to the either self-declared reporting level or adds a (+) if the report has been submitted to the GRI and has been declared as GRI checked.
Reporting Levels Assessments	No	Yes, GC “Notable COP program”.	Yes, GRI Application Level. Permits reporting companies to either self-declare their reporting level or apply with GRI to get their reporting level checked and being approved (checked) by GRI.
Members according to information provided by the Associations as of Jan 14th, 2013.	78 signatories and 1 associate member.	10,000+ signatory members since inception in 2003 but not necessarily all are currently reporting members.	2,047 organizations reported according to the GRI database in 2012. The entire database itself features 4,868 organizations. There are more than 600 organizational stakeholders who are members of the GRI in 2012.

Source: Authors, based on information provided by the three association’s websites retrieved in July, 2013.

Table 2: Matrix of CSR Reporting Activity by Firms (2000 to 2004)

	Frequency of A-Rated CSR Standards		
Frequency of D-Rated CSR Standards		Low	High
	Low	Bravo, Delta	Charlie
	High	Alpha, Echo	

Table 3: Matrix of CSR Reporting Activity by Firms (2010 to 2012)

	Frequency of A-Rated CSR Standards		
Frequency of D-Rated CSR Standards		Low	High
	Low	Alpha, Delta	Bravo, Charlie
	High	Echo	

References

- Adams, C. & Ambika, Z., "Corporate social responsibility: why business should act responsibly and be accountable," *Australian Accounting Review*, 14(34), (2004), 10-17.
- Adams, C.A., "Internal organizational factors influencing corporate social and ethical reporting: beyond current theorizing," *Accounting, Auditing & Accountability Journal*, 15(7), (2002), 223-50.
- Arevalo, J.A., "Low level of commitment phobia: An empirical study of the UN Global Compact's CSR movement," Doctoral dissertation, Rutgers, the State University of New Jersey, Newark, NJ, 2008.
- Argyris, C. & Schoen, D., *Organizational learning: A theory of action perspective*. Reading, MA: Addison Wesley, 1978.
- Argyris, C. 1982. *Reasoning, Learning, and action: Individual and organizational*. San Francisco, CA: Jossey-Bass, 1982.
- Bayoud, N. S., Kavanagh, M., & Slaughter, G., "An empirical study of the relationship between corporate social responsibility disclosure and organizational performance: Evidence from Libya," *Global Conference on Business & Finance Proceedings*, 7(2), (2012), 12-26.
- Bernhut, S., "Corporate social responsibility, with Pratima Bansal," *Ivey Business Journal*, 66(4), (2002), 18-19.
- Canada, Parliament. House of Commons. "An Act to Establish the Financial Consumer Agency of Canada, and to Amend Certain Acts in Relation to Financial Institutions." Bill C-8, 37th Parliament, 1st Session, 2001. Ottawa: Ministry of Finance, 2001. (1st Reading, February 7, 2001).
- Canada, Public accountability statements (banks, insurance companies, trust and loan companies) regulations. Statutory Regulation 2002—133. Government of Canada. Ottawa: Minister of Justice, 2002.
- Chapple, W. & Moon, J., "Corporate social responsibility (CSR) in Asia: A seven-country study of CSR web site reporting," *Business and Society*, 44(4), (2005), 415-441.
- Cuesta-González, M. & Muñoz-Torres, M. J. & Fernández-Izquierdo, M. A., "Analysis of Social Performance in the Spanish Financial Industry Through Public Data: A Proposal," *Journal of Business Ethics*, 69(3), (2006), 289-304.
- Cuganesan, S., & Khan, H., "Assessing the reporting of stakeholder reciprocity in the Australian banking industry," *Journal of HRCA: Human Resource Costing & Accounting*, 12(2), (2008), 85-101.

DiMaggio, P. & Powell, W., "The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields," *American Sociological Review*, 48, (1983), 147-160.

Drori, G. & Meyer, J., & Hwang, H. (Eds.), *Globalization and Organization: World Society and Organizational Change*. Oxford: Oxford University Press, 2006.

Eisenhardt, K. M., "Building theories from case study research," *Academy of Management Review*, 14(4), (1989), 532-532.

Equator Principles [web page] The equator principles. Equator Principles. June 2006, http://www.equator-principles.com/resources/equator_principles.pdf, [Accessed 12 Feb 2012].

Equator Principles. 2011. [web page] Guidance note on equator principles implementation reporting, Equator Principles, 2011. http://equator-principles.com/resources/ep_implementation_reporting_guidance_note.pdf. [Accessed 15 Mar 2012].

Equator Principles. 2013. [web page] Members and reporting. Equator Principles, 2013. <http://www.equator-principles.com/index.php/members-reporting>. [Accessed 14 Jan 2013].

GRI. 2011a. RG: Sustainability reporting guidelines 3.1. Global Reporting Initiative. Amsterdam, Netherlands: GRI, 2011.

GRI. 2011b. Sustainability reporting guidelines & financial services sector supplement. Global Reporting Initiative. Amsterdam, Netherlands: GRI, 2011.

GRI. 2011c. What is GRI? Global Reporting Initiative, 2011 <https://www.globalreporting.org/information/about-gri/what-is-GRI/Pages/default.aspx>. [Accessed 11 Dec 2011].

Idowu, S. O., & Towler, B. A., "A comparative study of the contents of corporate social responsibility reports of UK companies," *Management of Environmental Quality*, 15(4), (2004), 420-437.

Joyner, B.E. and Payne, D., "Evaluation and implementation: a study of values, business ethics and corporate social responsibility," *Journal of Business Ethics*, 41(4), (2002), 297-311.

Khan, Md Habib-Uz-Zaman & Halabi, A.K. & Martin, S., "Corporate social responsibility (CSR) reporting: a study of selected banking companies in Bangladesh," *Social Responsibility Journal*, 5(3), (2009), 344-357.

Khan, Md Habib-Uz-Zaman, "The effect of corporate governance elements on corporate social responsibility (CSR) reporting: Empirical evidence from private commercial banks of Bangladesh," *International Journal of Law and Management*. 52(2), (2010), 82-109.

King, A. "How to get started in corporate social responsibility," *Financial Management*, (Oct 2002),5.

- Margolis, J.D. & Walsh, J.P., "Misery loves companies: rethinking social initiatives by business," *Administrative Science Quarterly*, 48(2), (2003), 268-305.
- Marx, T.G., "Corporate social performance reporting," *Public Relations Quarterly*, 37(4), (1992/1993), 38-44.
- Miles, M.B. & Huberman, A.M., *Qualitative data analysis*. London: Sage, 1994.
- Mondal, A., & Ghosh, S. K., "Intellectual capital and financial performance of Indian banks," *Journal of Intellectual Capital*, 13(4), (2012), 515-530
- Nikolaeva, R., & Bicho, M., "The role of institutional and reputational factors in the voluntary adoption of corporate social responsibility reporting standards," *Academy of Marketing Science Journal*, 39(1), (2011), 136-157.
- Oliver, C., "Strategic responses to institutional processes," *Academy of Management Review*, 16(1), (1991), 145-179.
- Pérez, A., & Rodríguez, D. B., "The role of CSR in the corporate identity of banking service providers," *Journal of Business Ethics*, 108(2), (2012), 145-166
- Pike, A., "When life is more than just a lottery: Management social reporting," *Financial Times*, 11(9), (2000), 31-32.
- Reynolds, M. & Yuthas, K., "Moral discourse and corporate social responsibility reporting," *Journal of Business Ethics*, 78(1-2), (2008), 47-64.
- Rugman, A. & D'Cruz, J., *Multinationals as flagship firms: Regional business networks*, Oxford: Oxford University Press, 2000.
- Simms, J., "Business: Corporate social responsibility - You know it makes sense," *Accountancy*, 130(1311), (2002), 48.
- Tapscott, D. & Ticoll, D., *The naked corporation: How the age of transparency will revolutionize business*, Toronto: Penguin, 2003.
- United Nations. [web page] Secretary-General proposes global compact on human rights, labour, environment, in address to World Economic Forum in Davos. United Nations, <http://www.un.org/News/Press/docs/1999/19990201.sgsm6881.html>. [Accessed 12 Dec 2011].
- UNGC. [web page] Overview of the UN Global Compact. United Nations, (2011), <http://www.unglobalcompact.org/AboutTheGC/index.html>. [Accessed 12 Jan 2013].
- Vurro, C. & Perrini, F., "Making the most of corporate social responsibility reporting: Disclosure structure and its impact on performance," *Corporate Governance*, 11(4), (2011), 459-474.
- Weber, R. P., *Basic content analysis*, London: Sage, 1990.

ASB 2013
Antigonish, Nova Scotia

Jun Zhou
Dalhousie University

Laurence Booth
University of Toronto

Christos Ntantamis
Mount Allison University

FINANCIAL CONSTRAINTS, R&D INVESTMENT, AND THE VALUE OF CASH

This Version: August 2013

This paper investigates whether the relationship between financial constraints and the marginal value of corporate cash holdings varies across firms with and without R&D investment. We find that among firms with positive R&D investment the marginal value of cash is higher in financially constrained firms than unconstrained ones, whereas this difference is not significant among firms without R&D investment.

Key words: Value of Cash, Research and Development, Financial Constraints

JEL classification: G31, G32

1. Introduction

One important area of research in corporate finance is the impact of financial constraints on investment and financial policies. Capital market frictions increase the cost of external financing. Firms with great frictions in raising external financing (i.e. financially constrained) suffer from higher financing costs and may even miss some value-enhancing growth opportunities. As investment and financial policies are naturally intertwined, firms may adjust their financial management to attenuate the negative impacts of financial constraints. Liquidity management, in particular corporate cash holdings, has been studied by several recent papers. Almeida, Campello, and Weisbach (2004) find that financially constrained firms have higher propensity to save cash out of cash flows. Faulkender and Wang (2006) find that cash holdings are more valuable for financially constrained firms than for unconstrained firms. Together, these studies show that hoarding cash is a value-increasing activity for constrained firms.

Our study aims to extend this literature by examining whether the relationship between financial constraints and value of cash holdings would vary across firms with different involvement in R&D activity. This question is important for several reasons. First, the literature indicates a

strong link among financial constraints, investment, and financial policies, so it is interesting to consider these three components jointly. Further, relative to physical investments measured by capital expenditure, R&D investment has some unique features, including high information asymmetry, lack of collateral and high adjustment costs, that indicate a more important role of cash holdings on investments. In particular, Brown and Petersen (2011) provide empirical evidence that financially constrained firms use their cash holdings to help smooth R&D in spite of volatile and limited sources of external and internal financing.

We apply an extended version of the empirical methodology designed by Faulkender and Wang (2006) to a sample of publicly listed manufacturing firms in the United States over the period of 1972-2007. We split the sample into firms with and without R&D investment and examine the relationship between financial constrained and the value of cash in these subgroups respectively. For firms with positive R&D investment, we find that the value of an additional dollar of cash is higher for financially constrained firm than for unconstrained firms, and this relationship becomes economically and statistically significant in the second period (1984-1995) and third period (1996-2007). On the other hand, for firms without R&D investment, the difference is not significant for all three periods. These findings are robust to alternative measures used to classify firms into financially constrained and unconstrained firms.

This study is related to several strands of literature. First, this paper is directly linked to Brown and Petersen (2011). Using dynamic R&D models, Brown and Petersen (2011) find empirical evidence that firms most likely to face financial constraints rely on cash holdings to smooth R&D investment, while financially unconstrained firms do not. As maintaining a smooth path of R&D investment is important due to the high cost of adjustment, their empirical finding implies that among firms with R&D investment cash holdings should be more valuable for financially constrained ones than for unconstrained firms. This implication is directly tested and supported by our paper. Second, this paper also contributes to the literature on financial constraints. Although existing studies have examined the relationship between financial constraints and the value of cash holdings, our paper moves one step further, investigating whether this relationship varies according to the nature of investment.

The remainder of the paper is organized as follows. Section 2 reviews the related literature and develops testing hypotheses. Section 3 describes the empirical methodology and data. Empirical findings are presented in Section 4. Section 5 concludes and discusses potential policy implications.

2. Literature Review and Hypothesis Development

A large literature in corporate finance examines how financial constraints, i.e. the frictions in access to external capital markets, influence corporate investments and financial policies. The seminal paper, Fazzari et al. (1998) focuses on investments, and proposes that when a firm is more likely to be financially constrained its investment will vary with the availability of internal funds, instead of just with the availability of positive NPV projects. Hence, the investment to cash flow sensitivity will increase with the degree of financial constraints. Although their work has

been challenged on both theoretical grounds (Kaplan and Zingales, 1997; Povel and Raith, 2001) and empirical grounds (Kaplan and Zingales, 1997; Cleary, 1999; Erickson and Whited, 2000), their insight on the importance of financial constraints to corporate finance decisions is far-reaching.

On investigating the value of corporate cash holdings, Faulkender and Wang (2006) find that an additional dollar of cash holdings is more valuable for shareholders in financially constrained firms than for unconstrained firms. This finding can be attributed to two reasons. First, holding cash can help firms avoid the high cost of external financing. Second and more important, cash reserves allow firms to take value-increasing investment opportunities when other sources of financing are costly or limited. Denis and Sibilkov (2011) examine whether the difference in the value of cash between constrained and unconstrained firms is related to investment policy and find empirical evidence to support the second reason.

Although the aforementioned studies indicate a strong link among financial constraints, investment, and value of cash holdings, their focus is on physical investment as measured by capital expenditure, ignoring research and development (R&D), another important type of investment. Especially, Brown, Fazzari and Petersen (2009) show that R&D investment has increased significantly in the United States since 1980s and has become the principal investment for many public firms.

Compared to physical investment, R&D investment has some unique features (see Hall (2002) for a more detailed literature review). First, the nature of R&D prevents outsiders from making accurate assessment about its likelihood of success and the value of the innovative project. Firms have less incentive to reduce this high information asymmetry due to strategic concerns that competitors can imitate the innovative ideas. The difficulty of access risk characteristics and default probabilities, along with the lack of collateral, implies the difficulty to finance R&D through debt markets. Meanwhile, the high information asymmetry indicates high cost of equity financing as la Myers and Majluf (1984). Second, a sizable portion of R&D spending is the salaries of highly trained scientists and engineers, and the knowledge accumulated through R&D investment is embedded in the human capital of these R&D workers. Hence, R&D investment involves high adjustment costs as the adjustment tends to involve layoffs and rehires. Rehiring involves high training costs to build up proper knowledge. Further, knowing proprietary information, fired workers are able to transmit valuable knowledge to competitors. To avoid or minimize these high adjustment costs, firms tend to maintain a smooth level of R&D investment.

Due to these unique features of R&D investment, the relationship between financial constraints and the value of cash holdings may be different between firms with and without R&D investment. Hence, we propose the following two hypotheses to test in this paper.

Hypothesis 1: Among firms with R&D spending, cash holding is more valuable for financially constrained firms than for financially unconstrained firms.

This hypothesis is based on the following three reasons. For firms with R&D spending, cash holdings is more valuable for financially constrained firms as it helps them avoid the high cost

and limited access of external financing that is mainly driven by information asymmetry. Furthermore, R&D projects tend to have high-risk and high-return. Cash holding is more valuable for financially constrained firms as it helps them avoid missing potentially value-increasing projects. Last but not least, Brown and Petersen (2011) find empirical evidence that financially constrained firms rely extensively on cash holdings to smooth R&D, while financially unconstrained firms do not. Hence, cash holdings should be more valuable for financial constrained firms as it is essential for R&D smoothing.

Compared with firms relying on R&D investment, the level of information asymmetry between insiders and outside investors is less severe for firms without R&D investment, so external financing may be less costly. Further, the growth opportunities of these firms are based on physical investments, which can be used as collateral for debt financing in the future. Based on these, we reach the following hypothesis.

Hypothesis 2: Compared to firms with R&D spending, the relation between financial constraints and the value of cash is weaker among firms without R&D spending.

3. Methodology and Data

3.1 Empirical Methodology

To examine the difference in the marginal value of cash between positive-R&D firms and zero-R&D firms, we build on the methodology proposed by Faulkender and Wang (2006) with a slight modification.

$$\begin{aligned}
r_{i,t} - R_{i,t}^B = & \gamma_0 + \gamma_1 * \text{Constrained} * \frac{\Delta \text{Cashholdings}_{i,t}}{M_{i,t-1}} + \gamma_2 * \text{Constrained} + \gamma_3 * \frac{\Delta \text{Cashholdings}_{i,t}}{M_{i,t-1}} \\
& + \gamma_4 * \frac{\Delta \text{Earnings}_{i,t}}{M_{i,t-1}} + \gamma_5 * \frac{\Delta \text{NetAssets}_{i,t}}{M_{i,t-1}} + \gamma_6 * \frac{\Delta R \& D_{i,t}}{M_{i,t-1}} + \gamma_7 * \frac{\Delta \text{Interests}_{i,t}}{M_{i,t-1}} + \gamma_8 * \frac{\Delta \text{Dividends}_{i,t}}{M_{i,t-1}} \\
& + \gamma_9 * \frac{\text{Cashholdings}_{i,t-1}}{M_{i,t-1}} + \gamma_{10} * \text{Leverage}_{i,t} + \gamma_{11} * \frac{\text{NetFinancing}_{i,t}}{M_{i,t-1}} \\
& + \gamma_{12} * \frac{\text{Cashholdings}_{i,t-1}}{M_{i,t-1}} * \frac{\Delta \text{Cashholdings}_{i,t}}{M_{i,t-1}} + \gamma_{13} * \text{Leverage}_{i,t} * \frac{\Delta \text{Cashholdings}_{i,t}}{M_{i,t-1}} + \varepsilon_{i,t}
\end{aligned} \tag{1}$$

This methodology is essentially a long-run event study, investigating the impact of the change in corporate cash holdings on a firm's valuation over the event window of a fiscal year. The dependent variable is the excess return for firm i over fiscal year t , computed as stock i 's return minus the return of stock i 's benchmark portfolio over fiscal year t . The benchmark portfolios are the twenty-five Fama and French portfolios, formed by an independent sort of stocks on their size and book-to-market characteristics. The benchmark returns are used to control for the variation in the discount rate due to its sensitivity to time-varying risk factors.

Besides the change in cash holdings, the independent variables also include the changes in the firm's profitability (measured by earnings before extraordinary items), investment policy (including net assets and R&D expenditures), and financial policy (including interest expense, dividends, leverage, and net financing). Except for leverage, all the independent variables are scaled by the lagged market value of equity ($M_{i,t-1}$). Since the dependent variable can be rewritten as the change in the market valuation of equity divide by the lagged market value, the coefficient estimates can be interpreted directly as the value that shareholders place on a one-dollar change in the related independent variables. The independent variables also include the interaction terms of the change in cash with lagged cash and leverage, since the marginal value of cash to shareholders may decrease with the firm's existing cash position and the level of leverage.

To investigate the difference in the value of cash between financially constrained and unconstrained firms, we include an indicator variable (*Constrained*) equal to 1 for firms that are more likely to be financially constrained, and an interaction term between the change in cash and this indicator variable. The coefficient estimate on this interaction term ($\hat{\gamma}_1$) directly represents the difference in the marginal value of cash between financially constrained and unconstrained firms. The regression model is estimated with industry and year fixed effects. As suggested by Petersen (2009), robust standard errors clustered by firm are used to account for within-firm correlation of the error terms, i.e. the observations are assumed to be independent across firms, but not within firms.

3.2 Data

The base sample contains all U.S. publicly traded firms in the CRSP-Compustat merged database (Fundamental Annual) over the period of 1972-2007. We only keep manufacturing firms (SIC 2000-3999) in this study as most corporate R&D occurs in this sector and the literature on financial constraints tends to focus on this sector. Firm-year observations are deleted if there is missing value for any of the variables in equation (1). All the variables are winsorized at the one percent tails to reduce the effect of outliers. The sample screening leaves an unbalanced panel of 53,184 firm-year observations.

We divide the sample into three sample periods of equal length: 1972-1983, 1984-1995, and 1996-2007. According to whether a firm reports positive R&D in each given sample period, we divide the firms into 'positive R&D' and 'no R&D' subsamples. We use firm age, the number of years since the firm appeared on Compustat with a non-missing stock price, as our primary criterion to classify financially constrained and unconstrained firms. Firm age has been associated with financial constraints in the literature (Rauh 2006; Fee et al. 2009). Hadlock and Pierce (2010) collect detailed qualitative information on financial constraints from financial filings. They find that firm age and size are two most useful predictors of financing constraint level based on qualitative information. Moreover, these two measures are less endogenous than other measures. As younger firms are more likely to be financially constrained, a firm is classified as 'constrained' if its age in a given year is less than or equal to ten years.

Table 1 provides the summary statistics for each subsample. For firms with positive R&D spending, young firms tend to hold more cash than mature firms. The difference is very

significant in the second and third periods, with both mean and median ratios of cash to total assets significantly higher for young firms. This is consistent with the literature that firms classified as financially constrained tend to hold more cash (Almeida et al. 2004; Faulkender and Wang, 2006; Denis and Sibilkov, 2011). We also notice that young firms' R&D investment is more intensive than mature firms in all three periods. Moreover, both mean and median R&D investment ratios increase considerably for young firms, while the increase is less significant for mature firms. The sample of firms without R&D investment is smaller than the sample of firms with positive R&D. Compared to positive R&D firms, firms without R&D tend to hold less cash and the difference, in both mean and median, has widened over time. Moreover, mean and median cash ratios are very similar for young and mature firms in all three periods.

Table 1: Summary Statistics

			1972-1983		1984-1995		1996-2007		
			Young	Mature	Young	Mature	Young	Mature	
Zero R&D firms	Cash/TA	Mean	0.084	0.083	0.091	0.089	0.080	0.087	
		Median	0.048	0.048	0.035	0.035	0.029	0.032	
	RD/TA	Mean	0	0	0	0	0	0	
		Median	0	0	0	0	0	0	
	Observations			2072	2742	2320	2933	1516	2555
	Positive R&D firms	Cash/TA	Mean	0.100	0.081	0.227	0.108	0.312	0.175
Median			0.051	0.052	0.138	0.059	0.235	0.096	
RD/TA		Mean	0.132	0.079	0.189	0.079	0.192	0.087	
		Median	0.077	0.052	0.111	0.046	0.102	0.042	
Observations			4565	8038	6397	7188	4892	7966	

4. Empirical Results and Analysis

4.1 Young Firms vs. Mature Firms

Table 2 and Table 3 report the estimates of regression model (1) when firm age is used to classify firms into financially constrained and unconstrained groups.

In Table 2, the first column contains the regression result for the whole sample. The coefficient estimate of the interaction term between constrained dummy and the change in cash is positive and statistically significant at 1% level. This indicates that shareholders put a higher marginal value on the cash holdings of young firms than on the cash of mature firms. This result is consistent with previous studies that show cash is more valuable for financially constrained firms than for unconstrained firms (Faulkender and Wang, 2006; Denis and Sibilkov, 2010).

Table 2: Firm Age and the Value of Cash

Constrained according to Age	1972-2007	1972-1983	1984-1995	1996-2007
Constrained*Change in cash	0.154	0.036	0.246	0.211
	(0.001)	(0.570)	(0.002)	(0.027)
Constrained	-0.020	-0.006	-0.025	-0.031
	(0.000)	(0.444)	(0.001)	(0.001)
Change in cash	1.530	1.128	1.412	1.952
	(0.000)	(0.000)	(0.000)	(0.000)
Change in earnings	0.508	0.517	0.488	0.517
	(0.000)	(0.000)	(0.000)	(0.000)
Change in net assets	0.212	0.162	0.248	0.246
	(0.000)	(0.000)	(0.000)	(0.000)
Change in R&D	1.109	1.400	1.270	0.687
	(0.000)	(0.000)	(0.000)	(0.002)
Change in interests	-1.671	-1.619	-1.619	-1.868
	(0.000)	(0.000)	(0.000)	(0.000)
Change in dividends	3.252	4.293	3.509	1.613
	(0.000)	(0.000)	(0.000)	(0.013)
Lagged cash	0.276	0.260	0.256	0.328
	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.502	-0.415	-0.535	-0.564
	(0.000)	(0.000)	(0.000)	(0.000)
Net financing	0.031	0.030	0.049	-0.032
	(0.054)	(0.185)	(0.104)	(0.387)
Change in cash*lagged cash	-0.914	-0.682	-0.909	-1.292
	(0.000)	(0.000)	(0.000)	(0.000)
Change in cash*leverage	-1.596	-1.098	-1.353	-1.780
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.231	0.190	0.104	0.128
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	53,184	17,417	18,838	16,929
Adj. R-squared	0.220	0.242	0.220	0.219

This table reports the results of regression model (1) for the full sample period of 1972-2007 and three sub-periods of equal length. In this table, we use firm age to classify firms into financially constrained and unconstrained groups. Constrained is an indicator variable that is equal to 1 if a firm's age is less than or equal to ten years. All specifications include year fixed effects and industry fixed effects. The standard errors are adjusted for clustering on firms. P-values are reported in parentheses. For each regression, adjusted R² is reported.

To examine whether this relation varies over time, we conduct the test in each period and report the results in Columns [2]-[4]. For all three periods, the coefficient estimate of the interaction term between constrained dummy and the change in cash is positive, but it is statistically and economically significant for the second period (1984-1995) and third period (1996-2007). For

economical interpretation, the coefficient estimates of this interaction term imply that the marginal value of cash is about 25 cents higher for young firms than for mature firms during 1984-1995, and the difference is about 21 cents during 1996-2007. These findings are related to the new listing effect proposed in Fama and French (2004). Firms that went public in the 1980s and 1990s tend to have more growth opportunities and are less profitable than earlier firms. This implies that young firms in the second and third periods may face more severe financial constraints. Therefore, the importance of cash holdings for young firms, as a means of internal financing, is acknowledged by investors through higher market valuation.

To examine whether the relation between financial constraints and the value of cash varies across firms with and without R&D investment, Table 3 provides estimates of regression (1) for these two groups respectively in each of the three periods. We focus our analysis on the second and third time periods as the new listing effect leads to a stronger contrast between young firms and mature firms in terms of the likelihood to be financially constrained.

For firms with positive R&D spending, the coefficient estimates on the interaction term between constrained dummy and the change in cash are positive (0.293 for 1984-1995 period and 0.284 for 1996-2007 period) and statistically significant at 1% level. This finding supports our first hypothesis. The value of coefficient estimates on this interaction term implies that among those firms with positive R&D spending the marginal value of cash is about 28 to 29 cents higher in young firms than in mature firms. This finding is directly related to Brown and Petersen (2011), that young firms use cash reserves to smooth their R&D investment while mature firms can rely on other sources for financing. The summary statistics in Table 1 shows that R&D investment is more intensive in young firms than mature firms. Hence, for those firms with R&D investment, cash should be more valuable for young firms as it is essential to their investment and growth by avoiding cutting R&D and missing growth opportunities. In this sense, our analysis supports the view that a policy of great cash retention by young firms with R&D investment tends to be a value-enhancing policy.

On the other hand, the coefficient estimates on the interaction term between constrained dummy and the change in cash are statistically insignificant for firms without R&D spending. Compared to the statistically and economically significant relationship between financial constraints and the value of cash among firms with R&D investment, this finding here implies a weak relationship between the value of cash and financial constraints (measured by firm age here) when firms are not involved in R&D investment. This finding supports our second hypothesis.

Table 3: Firm Age, R&D, and the Value of Cash

Constrained according to Age	1972-1983		1984-1995		1996-2007	
	Positive R&D	Zero R&D	Positive R&D	Zero R&D	Positive R&D	Zero R&D
Constrained*Change in cash	0.087 (0.281)	-0.015 (0.884)	0.293 (0.001)	0.104 (0.483)	0.284 (0.009)	-0.102 (0.619)
Constrained	-0.005 (0.600)	-0.010 (0.414)	-0.036 (0.000)	0.000 (0.969)	-0.044 (0.000)	0.008 (0.622)
Change in cash	1.151 (0.000)	0.970 (0.000)	1.484 (0.000)	0.982 (0.000)	2.021 (0.000)	1.558 (0.000)
Change in earnings	0.584 (0.000)	0.412 (0.000)	0.519 (0.000)	0.425 (0.000)	0.533 (0.000)	0.487 (0.000)
Change in net assets	0.181 (0.000)	0.134 (0.000)	0.248 (0.000)	0.244 (0.000)	0.283 (0.000)	0.189 (0.000)
Change in R&D	1.407 (0.000)	1.227 (0.362)	1.313 (0.000)	2.158 (0.338)	0.695 (0.003)	2.369 (0.314)
Change in interests	-1.975 (0.000)	-1.100 (0.000)	-1.701 (0.000)	-1.478 (0.000)	-1.939 (0.000)	-1.521 (0.000)
Change in dividends	3.689 (0.000)	5.306 (0.000)	2.746 (0.000)	4.697 (0.000)	0.426 (0.640)	3.306 (0.000)
Lagged cash	0.266 (0.000)	0.243 (0.000)	0.295 (0.000)	0.148 (0.013)	0.368 (0.000)	0.187 (0.003)
Leverage	-0.426 (0.000)	-0.390 (0.000)	-0.563 (0.000)	-0.498 (0.000)	-0.604 (0.000)	-0.529 (0.000)
Net financing	0.023 (0.434)	0.036 (0.323)	0.071 (0.076)	-0.004 (0.931)	-0.057 (0.214)	-0.030 (0.599)
Change in cash*lagged cash	-0.726 (0.000)	-0.547 (0.004)	-0.997 (0.000)	-0.604 (0.034)	-1.425 (0.000)	-0.753 (0.015)
Change in cash*leverage	-1.068 (0.000)	-0.982 (0.000)	-1.357 (0.000)	-0.806 (0.009)	-1.638 (0.000)	-1.572 (0.001)
Constant	0.172 (0.000)	0.232 (0.000)	0.043 (0.233)	0.168 (0.000)	0.100 (0.001)	0.175 (0.000)
Observations	12,603	4,814	13,585	5,253	12,858	4,071
Adj. R-squared	0.247	0.235	0.228	0.224	0.223	0.237

This table reports the results of regression model (1) for firms with and without R&D investment respectively during three sub-periods of equal length. In this table, we use firm age to classify firms into financially constrained and unconstrained groups. 'Constrained' is an indicator variable that is equal to 1 if a firm's age is less than or equal to ten years. All specifications include year fixed effects and industry fixed effects. The standard errors are adjusted for clustering on firms. P-values are reported in parentheses. For each regression, adjusted R² is reported.

4.2 Alternative Measures of Financial constraints

We further perform the analysis based on several alternative approaches for sorting firms into groups more or less likely to face financial constraints (Fazzari et al. 1988; Gilchrist and Himmelberg, 1995; Almeida et al. 2004; Faulkender and Wang, 2006; Rauh 2006; Fee et al. 2009; Dennis and Sibilkov, 2011; etc.).

- *Payout*. Compared to financially constrained firms, unconstrained firms are more likely to distribute cash to their investors. We assign those firms with positive cash payout, defined as cash dividends plus stock buybacks minus stock issuances, to the financially unconstrained group. Firms with zero or negative cash payout are classified as constrained.
- *Dividend payment*. Although share repurchase has become a popular way to distribute cash to shareholders, it is used by some firms to management earnings, offset stock option dilution, etc. (Brav, et al., 2005) Hence, using payout to classify firms may introduce some noise. To address this issue, we classify firms according to its annual cash dividend payment. Firms without cash dividends are assigned to financially constrained group, while firms with positive cash dividends are classified as unconstrained.
- *Firm size*. In general smaller firms are younger and less known in capital markets, so they tend to suffer more from capital market imperfections. We split firms into large and small size groups based on their sales at the end of the previous fiscal year. Those firms in the top three deciles of the firm size distribution are classified as the unconstrained group. Firms in the other deciles are classified as constrained.
- *Bond ratings*. Bond ratings stand for the market's assessment of a firm's credit quality since a firm need to reach a certain quality before the risk associated with information asymmetry is low enough to make bond issuance feasible. We split firms based on whether or not a bond rating (S&P long-term senior debt rating) is reported in Compustat and categorize those firms that did not have their public debt rated as financially constrained.

Table 4: Correlation of Various Classifications of Financial Constraints

	Financial constraints criteria			
	Age	Dividend	Payout	Size
Dividend	0.374 (0.000)			
Payout	0.321 (0.000)	0.741 (0.000)		
Size	0.351 (0.000)	0.433 (0.000)	0.348 (0.000)	
Bond Rating	0.247 (0.000)	0.339 (0.000)	0.264 (0.000)	0.630 (0.000)

Correlation coefficient estimates are reported with p-values in parentheses below.

Table 4 reports Pearson's correlation coefficients for these four alternative measures of financial constraints, as well as firm age. All correlation coefficient estimates are positive and statistically significant at 1% level, indicating that these measures capture some common information about

financial constraints. The correlation coefficient estimates range from 0.247 to 0.741, implying that each measure may contain some unique information as well. Therefore, we conduct further analysis using each of these alternative criteria to classify firms as financially constrained or unconstrained.

Table 5: Alternative Financial Constraints Measures and the Value of Cash

1984-1995								
	Payout		Dividend		Size		Bond ratings	
	Positive R&D	Zero R&D	Positive R&D	Zero R&D	Positive R&D	Zero R&D	Positive R&D	Zero R&D
Constrained*Change in cash	0.355	-0.106	0.439	-0.038	0.320	0.182	0.210	0.168
	(0.000)	(0.436)	(0.000)	(0.805)	(0.002)	(0.191)	(0.067)	(0.315)
Constrained	0.039	0.035	-0.002	0.039	-0.022	-0.002	-0.044	-0.003
	(0.000)	(0.015)	(0.823)	(0.007)	(0.013)	(0.893)	(0.000)	(0.807)
Change in cash	1.411	1.106	1.334	1.062	1.416	0.888	1.526	0.928
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Change in cash*lagged cash	-0.977	-0.595	-1.055	-0.593	-1.036	-0.609	-1.029	-0.584
	(0.000)	(0.031)	(0.000)	(0.033)	(0.000)	(0.031)	(0.000)	(0.047)
Change in cash*leverage	-1.522	-0.823	-1.516	-0.833	-1.443	-0.788	-1.539	-0.809
	(0.000)	(0.007)	(0.000)	(0.007)	(0.000)	(0.010)	(0.000)	(0.012)
Observations	13,585	5,253	13,585	5,253	13,585	5,253	12,701	4,881
Adj. R-squared	0.228	0.225	0.227	0.225	0.227	0.224	0.227	0.225
1996-2007								
	Payout		Dividend		Size		Bond ratings	
	Positive R&D	Zero R&D	Positive R&D	Zero R&D	Positive R&D	Zero R&D	Positive R&D	Zero R&D
Constrained*Change in cash	0.148	0.028	0.236	-0.314	0.331	-0.353	0.249	-0.054
	(0.214)	(0.883)	(0.090)	(0.186)	(0.027)	(0.104)	(0.091)	(0.796)
Constrained	0.029	0.045	-0.053	0.019	-0.104	-0.092	-0.125	-0.091
	(0.005)	(0.015)	(0.000)	(0.268)	(0.000)	(0.000)	(0.000)	(0.000)
Change in cash	2.050	1.471	1.966	1.700	1.868	1.810	1.926	1.548
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Change in cash*lagged cash	-1.360	-0.714	-1.369	-0.689	-1.372	-0.673	-1.333	-0.718
	(0.000)	(0.023)	(0.000)	(0.029)	(0.000)	(0.036)	(0.000)	(0.022)
Change in cash*leverage	-1.730	-1.543	-1.735	-1.441	-1.642	-1.738	-1.664	-1.620
	(0.000)	(0.002)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.001)
Observations	12,858	4,071	12,858	4,071	12,858	4,071	12,858	4,071
Adj. R-squared	0.221	0.238	0.222	0.238	0.226	0.244	0.227	0.241

This table reports the results of regression model (1) for firms with and without R&D investment respectively during three sub-periods of equal length. We use four alternative criteria, payout, dividend, size and bond ratings, to classify firms into financially constrained and unconstrained groups. See the text for definitions of criteria used to classify firm into financially constrained and

unconstrained groups. Constrained is an indicator variable that is equal to 1 if a firm is classified as more likely to be financially constrained according to a given criterion. All specifications include year fixed effects and industry fixed effects. The standard errors are adjusted for clustering on firms. P-values are reported in parentheses. For each regression, adjusted R^2 is reported.

We estimate the regression model (1) for each subsample respectively. However, to conserve space, we only report in Table 5 coefficient estimates on those items that are related to cash holdings. The results for the second period (1984-1995) and the final period (1996-2007), based on each of these four alternative criteria, are reported in Table 5. The results are very similar to results from the young vs. mature firms split documented in Table 2. More specifically, for firms with positive R&D spending, the coefficient estimates for the interaction of the financial constraints dummy with the change in cash is positive and statistically significant for all alternative criteria of financial constraints. The only exception is for the 1996-2007 period when payout is used to measure financial constraint. The coefficient estimates imply that the marginal value of cash holdings is between 21 and 44 cents higher for financially constrained firms than for financially unconstrained firms in 1984-1995 period, or between 15 and 33 cents higher for financially constrained firms in 1996-2007 period. For firms without R&D investment, the coefficient estimate of the interaction term between financial constraints dummy variable and the change in cash is statistically insignificant. This indicates a weak relationship between financial constraints and the value of cash among zero R&D firms.

Overall, the results documented in Table 5 provide further support for our testing hypotheses as our results are robust to alternative approaches for sorting firms into groups more or less likely to face financial constraints.

5. Conclusion

This paper investigates the relationship between financial constraints and the value of cash holdings for firms with and without R&D investment respectively. We find that the value of cash is higher in financially constrained firms than unconstrained firms among firms with positive R&D, whereas this difference in the value of cash holdings across firms with different levels of financial constraints is not significant among firms without R&D investment.

The findings in our paper show that the importance of cash holdings for financially constrained firms varies according to the nature of investment. The unique features of R&D investment indicate a more important role of cash holdings in financial management, especially for financially constrained firms (Brown and Petersen, 2011). This paper shows that accumulating cash is a value-increasing activity for constrained firms with positive R&D. To some extent, our finding is also consistent with Bates, Kahle, and Stulz (2009). They identify a secular increase in corporate cash holdings over recent decades and attribute this partially to increasing R&D intensity. Our paper shows that cash holdings have increased sharply for firms with R&D investment, among which the level and the marginal value of cash holdings are always higher for constrained firms.

References:

- Almeida, H., M. Campello, and M. S. Weisbach. 2004. The Cash Flow Sensitivity of Cash. *Journal of Finance* 59:1777–804.
- Bates, T.W., K.M. Kahle, and R.M. Stulz, 2009. Why do U.S. Firms Hold So Much More Cash Than They Used To? *Journal of Finance* 64, 1985- 2021.
- Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely, 2005. Payout policy in the 21st century. *Journal of Financial Economics* 77, 483-527.
- Brown, J.R., Fazzari, S.M., Petersen, B.C., 2009. Financing Innovation and Growth: Cash Flow, External Equity and the 1990s R&D Boom. *Journal of Finance* 64,151–185.
- Brown, J. R., Petersen, B.C., 2011. Cash Holdings and R&D Smoothing. *Journal of Corporate Finance* 17, 694-709
- Cleary, S., 1999. The relationship between firm investment and financial status. *Journal of Finance* 54, 673–692.
- Denis, D.J. and V. Sibilkov. 2010. Financial Constraints, Investment, and the Value of Cash Holdings. *Review of Financial Studies*, 23 (1): 247-269.
- Erickson, Timothy, and Toni Whited, 2000. Measurement error and the relationship between investment and Q, *Journal of Political Economy* 108, 1027–1057.
- Fama, E.F. and K.R. French, 2004. New lists: fundamentals and survival rates. *Journal of Financial Economics* 73, 229–269.
- Faulkender, M., Wang, R., 2006. Corporate financial policy and the value of cash. *Journal of Finance* 61, 1957–1990.
- Fazzari, Steven, R. Glenn Hubbard, and Bruce Petersen, 1988. Financing constraints and corporate investment, *Brooking Papers on Economic Activity* 1, 141–195.
- Fee, C. E., C. J. Hadlock, and J. R. Pierce. 2009. Investment, Financing Constraints, and Internal Capital Markets: Evidence from the Advertising Expenditures of Multinational Firms. *Review of Financial Studies* 22:2361–92.
- Gilchrist, Simon, and Charles Himmelberg, 1995. Evidence on the role of cash flow for investment, *Journal of Monetary Economics* 36, 541–572.
- Hadlock, C.J. and J.R. Pierce. 2010. New evidence on measuring financial constraints: moving beyond the kz index. *Review of Financial Studies* 23(5), 1909-1940.
- Hall, B.H., 2002. The financing of research and development. *Oxford Review of Economic Policy* 18, 35-51.
- Kaplan, Steven, and Luigi Zingales, 1997. Do financing constraints explain why investment is correlated with cash flow? *Quarterly Journal of Economics* 112, 169–215.
- Myers, S., and N. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13:187–221.
- Povel, Paul, and Michael Raith, 2001. Optimal investment under financial constraints: The roles of internal funds and asymmetric information, Working paper, University of Chicago and University of Minnesota.
- Rauh, J. 2006. Investment and Financing Constraints: Evidence from the Funding of Corporate Pension Plans. *Journal of Finance* 61:33–72.

ASB 2013
Antigonish, Nova Scotia
September 27 – 29, 2013

Hui Zhu
Shannon School of Business
Cape Breton University

IMPLICATIONS OF LIMITED INVESTOR ATTENTION TO CUSTOMER-SUPPLIER INFORMATION TRANSFERS¹

This study focuses on the market reaction to information transfers from economically linked customers. I examine whether investors have limited attention with respect to the information contained in customer earnings announcements for suppliers. Using 1,083 unique customer-supplier relationships for the period 1983–2011, I find that the cumulative abnormal returns of a supplier surrounding and following linked customers' earnings announcements are positively related to earnings information of the customers. The results provide evidence that customer earnings announcements convey information for suppliers, suggesting that limited investor attention to customer-supplier information transfers from earnings announcements generates predictable returns across linked firms.

¹ I would like to thank anonymous reviewers of the 2013 Atlantic School of Business annual meeting, conference participants at the 2013 European Financial Association annual meeting and the 2013 Multinational Finance Society annual meeting for their valuable comments. I also acknowledge financial support from Cape Breton University RP Grant. All errors remain my own.

Introduction

This study aims to identify predictable returns by using ex ante economic links between customers and suppliers. Recent studies on the limited investor attention hypothesis,² which state that investors' limited attention to the arrival of new information causes return anomalies, show that investor inattention is more likely when a large number of same-day earnings announcements are made by other firms (Hirshleifer, Lim, and Teoh, 2009) or where there are a large number of Friday earnings announcements (Dellavigna and Pollet, 2009). Investor inattention is also more likely when publicly available information about economically linked firms is neglected (Cohen and Frazzini, 2008). More importantly, evidence on the limited attention to economically linked firms suggests that information diffuses from customers to suppliers, generating predictable returns across linked assets.

In this paper, I examine whether investors have limited attention with respect to the information contained in customer earnings announcements for suppliers. More specifically, I investigate the immediate responsiveness of a firm's abnormal returns surrounding the announcement dates of its linked customers as well as the delayed responsiveness of stock returns following the linked customers' earnings announcements. The disclosure of customer-supplier links between firms was required according to Statement of Financial Accounting Standards (SFAS) No. 14 before 1997 and based on SFAS No. 131 after 1997, and this information is available for public use. When news about a linked firm is released into the market, the stock price of the supplier firm should respond immediately to that news if investors consider these ex ante links. On the other hand, if investors pay limited attention to such links, the stock price of the supplier will react slowly to the linked firm's earnings news, and delayed abnormal returns can be expected. As a result, limited investor attention to a linked firm's announcements (i.e., a customer's unexpected earnings news³) leads to market underreactions.

Through the use of 1,083 unique customer-supplier relationships between 1983 and 2011, I find that the cumulative abnormal returns of a supplier surrounding and following linked customers' earnings announcements are positively related to unexpected earnings news of the customers. The findings indicate that there is a direct immediate relation between supplier returns and customer earnings surprises, and customer earnings surprises are positively related to post-customer earnings announcement supplier returns. Additionally, these results are robust to controlling the timing and order of the earnings announcement, same-industry effect, and the ratios of firm size, sales, and returns⁴ of linked customers to suppliers, and with respect to the delayed returns over different horizons. Because customer-supplier links between firms are typically associated with information transfer, the main results suggest that limited investor attention to the arrival of new information about economically linked firms generates abnormal stock returns.

The limited investor attention hypothesis by Cohen and Frazzini (2008) argues that customer returns predict future supplier's returns because investors pay limited attention to the customer-supplier link. To identify what role, if any, customer earnings surprises play in the customer-supplier return predictability, I further examine the effect of customer earnings surprises on the customer-supplier returns and find the positive enhancing effect of customer earnings surprises in the relationship between customer returns and supplier returns. The evidence supports the notion that investor's underreaction to the customer's earnings surprise leads to a drift in the customer's post-earnings announcement returns (Bernard and Thomas, 1989), which in turn is reflected in the supplier's returns due to the limited attention hypothesis. It also provides strong support that the customer's earnings surprise is a predominant source of the cross-firm return predictability.

² The limited attention hypothesis by Cohen and Frazzini (2008) states that stock prices underreact to firm-specific information that induces changes in the valuation of related firms, generating return predictability across assets. In particular, stock prices underreact to negative (positive) news involving related firms and, in turn, generate negative (positive) price drift.

³ Unexpected earnings news, unexpected earnings, earnings surprises, earnings news, and earnings-related information are exchangeable terms used in the paper.

⁴ Cohen and Frazzini (2008) provide evidence of return predictability across economically linked firms.

The paper contributes to existing literature in several ways. First, the paper adds to the growing stream of studies on the implications of limited attention for stock returns. Cohen and Frazzini (2008) examine how investors' limited attention to economically related firms leads to predictable future stock returns by testing "customer momentum," which is defined as a monthly strategy of buying firms whose customers had the most positive returns in the previous month and selling firms whose customers had the most negative returns in the previous month. If investors pay limited attention to the stock returns of economically linked firms, one would expect investors also to be inattentive to earnings-related information from such firms. Consequently, I hypothesize that market underreactions for suppliers are related to limited investor attention to earnings announcements by economically linked customers. I test this hypothesis by examining immediate (delayed) market reactions surrounding (following) earnings announcements by economically linked customers. Though most studies investigate market reactions around the time of a firm's own earnings announcements,⁵ I focus on market reactions around the time of earnings announcements by related firms because investors tend to ignore the publicly available link between suppliers and economically related customers. That is, investors are inattentive to customer-supplier links, and thus, stock returns are predictable.

Second, this study provides a new insight into information diffusion. The customer-supplier links between firms are longstanding public relationships. Thus, the earnings information released by customers is closely related to the earnings information for suppliers. Prior studies have shown that one firm's earnings news can be useful in updating earnings expectations for other firms in the industry.⁶ For instance, Ramnath (2002) examines intra-industry information diffusion by investigating the market reaction experienced by a firm that announces its earnings subsequent to the first announcing firm in the same industry when the earnings of the latter is unexpected. A recent study by Kovacs (2011) further shows that the firm's post-earnings announcement drift is driven by information diffusion from subsequent-announcing industry peers. If earnings information is transferred from other firms in the industry, one would also expect that investors perceive the earnings-related information from the economically related announcing-customers to be useful in updating their expectations for suppliers, and thus earnings information to be transferred from economically related firms. I find that the immediate and delayed returns of a supplier⁷ surrounding and following customers' earnings announcements are positively related to customers' unexpected earnings, confirming that customer earnings announcements convey information for suppliers.

Related Studies and Hypotheses

Limited Investor Attention

This paper builds on the finance literature on limited investor attention and its effects on financial markets. Recent empirical studies have related limited investor attention to asymmetric selling behavior (Barber and Odean, 2008), demographic shifts (Dellavigna and Pollet, 2007), and relevant information at the time of previous extraneous news (Huberman and Regev, 2001). Using share turnover as a proxy for investor attention, Hou, Peng, and Xiong (2009) show that price underreaction to earnings news is weak when investors are attentive, but the price drift caused by investors' overreaction is strong with investor attention. In the same vein, Loh (2010) finds that

⁵ Koch and Sun (2004) test announcement reactions around the firm's subsequent dividend changes, conditioning on the sign of the firm's unexpected earnings. Kovacs (2011) examines the effect of same-industry peers' earnings announcements on the post-earnings announcement drift. Cohen and Frazzini (2008) investigate market reactions to news about related firms but do not relate the phenomenon to the price response of suppliers around (after) the customer's earnings announcement.

⁶ Earnings-related information transfers in the industry examined in prior studies include, among others, Freeman and Tse (1992), Ramnath (2002), and Kovacs (2011).

⁷ On the basis of Regulation SFAS No. 131, it is the suppliers that need to report the identity of customers representing more than 10% of their total sales in interim financial reports issued to shareholders. As a result, customers are much larger firms than their suppliers. Thus, the customer is important to suppliers based on sales. However, there is no information available about how important the supplier is to the customer. Therefore, this study only focuses on the effects of customer earnings surprises on supplier returns.

investor inattention and the underreaction to stock recommendations lead to postrecommendation drift. In contrast, Da, Engelberg, and Gao (2011) use the Google search volume index (SVI) as a proxy for investor attention and show stronger price momentum among stocks with higher levels of SVI. Yuan (2012) finds that high market-wide attention generates heavy trading and price changes by analyzing the ability of market-wide attention-grabbing events, which are measured as record-breaking events for the Dow index and front-page articles about the stock market. Bae and Wang (2012) investigate whether the China-name affects investor attention and firm value and find that the returns of China-name stocks are, on average, more than 100% higher than those of non-China-name stocks. Bae and Wang attribute this phenomenon to increased investor attention to China-name stocks after controlling for alternative measures of investor attention, such as *Wall Street Journal* news coverage, abnormal trading volume, extreme past one-day returns, and the Google SVI. In addition, Gilbert, Kogan, Lochstoer, and Ozyildirim (2012) use the U.S. Leading Economic Index as a proxy for the stale information and find that investor inattention to the stale nature of information causes return anomalies.

The literature also discusses theoretical approaches to modeling limited investor attention. For instance, Merton (1987) suggests that higher expected stock returns are obtained from lesser-known stocks with smaller investors. Hong and Stein (1999) suggest that investor profit from trading on information is gradually transferred across the population if the information is helpful in predicting future outcomes. Hirshleifer and Teoh (2003) model how investors' inattention to accounting reports can lead to the misvaluation of stocks. Peng and Xiong (2006) demonstrate that investors are more likely to respond to market- and industry-wide information than they are to consider firm-specific information, which makes cross-sectional returns predictable. Finally, Peng (2005) develops a model in which the learning process for investors is optimally allocated when they have a limited capacity for information processing. Peng further predicts that mispricing is related to the speed with which investors process information about large or small firms.

Information Transfers

Several studies also focus on the role of information transfer in the prediction of future stock returns. For instance, Ramnath (2002) finds that the response to subsequent announcing firms around the first announcing date in the industry is positive when the earnings of the first announcing firm are unexpected. The underreaction, in turn, leads to predictable stock returns for subsequent announcers in the same industry. In the same vein, Hong, Torous, and Valkanov (2007) suggest that some industries predict future stock market returns. A recent study by Desir (2013) documents that managers of non-announcing firms are more likely to disclose good news to minimize the impact of negative information transfers from their industry competitors. Both Cohen and Frazzini (2008) and Menzly and Ozbas (2010) find that individual customer returns generate predictable future supplier returns. The former focuses more broadly on customer-supplier links. By contrast, the latter examines specific inter- and intra-industry relations.

In addition, Prokopczuk (2010) reports strong evidence that earnings news leads to a substantial contagion effect in the banking industry, but outside that industry, the magnitude of the contagion effect is positively related to the bank size and the size of the reporting news. Similarly, Jorion and Zhang (2012) investigate information transfer effects of bond rating downgrades and find evidence of a predominant contagion effect for investment-grade firms. A recent study by Cai, Song, and Walkling (2011) finds strong evidence that bidder abnormal returns are positively related to the degree of surprise associated with a bid announcement, and the prices of rival firms adjust at the time of an initial industry bid, suggesting the transfer of bid-related information through industry channels. Using the degree of accessibility of foreign investors to emerging stock markets as a proxy for investibility of foreign investments, Bae, Ozoguz, Tan, and Wirjanto (2012) find that greater investibility is associated with faster diffusion of global market information across stocks in emerging markets.

On the other hand, recent studies have shown that information transfers play an important role in the post-earnings announcement drift. Among others, Kovacs (2011) presents strong evidence that subsequent same-industry earnings announcements are related to a firm's post-earnings

announcement drift. Hou (2007) argues that industry information transfer from large firms to small firms contributes to the post-earnings announcement drift.

Hypothesis Development

My study is distinct from these other articles in that I analyze earnings-related information diffusion from customers to suppliers. That is, I test how a supplier's abnormal returns around the customers' earnings announcement date react to the unexpected earnings by linked customers. In a closely related paper, Cohen and Frazzini (2008) examine how limited investor attention to economically related firms leads to predictable future stock returns, whereas Ramnath (2002) investigates investor and analyst reactions to earnings announcements by related firms in the industry. A recent study by Kovacs (2011) further shows that the firm's post-earnings announcement drift is driven by information diffusion from subsequent-announcing firms in the same industry. If the corresponding price drift of the firm can be predicted based on earnings-related information from subsequent-announcing firms in the same industry, one would expect the stock price reactions of the firm surrounding the earnings report date of related firms to reflect this information. Therefore, investors can incorporate the information from linked customers' earnings announcements into their expectations for suppliers. In an efficient market, one would expect the immediate price responses for a supplier encompassing its linked customers' earnings announcements to reflect those announcements. Thus, the stock price responses of a supplier around its customers' announcements will be positively related to unexpected customer earnings. I therefore make the following hypothesis:

H_{1a}: The cumulative abnormal return of a supplier, surrounding the linked customers' announcement date, is positively related to linked customers' earnings news.

However, prior studies (e.g., Bernard and Thomas, 1989; Abarbanell and Bernard, 1992; Ramnath, 2002) have shown that investors cannot completely adjust their earnings expectations for announcing firms and that this dynamic leads to predictable stock returns (see the relationship in arrow #1 in Figure 1). More specifically, investor underreactions to first announcer's news yield predictable stock returns for subsequent announcers in the same industry (Ramnath, 2002). If investors are inattentive to earnings-related information from economically related customers, they cannot immediately react as strongly to the earnings-related information. Therefore, the corresponding drift in prices, i.e., the abnormal returns of a supplier cumulated after the customers' earnings announcement date, would be predictable even after the customers' actual earnings are released to the market. Thus, one would expect that the stock returns of a supplier cumulated after the customers' earnings announcement date will be positively related to unexpected customer earnings. It follows that:

H_{1b}: The abnormal return of a supplier, cumulated after the linked customers' announcement date, is positively related to linked customers' earnings news.

The discussion above suggests that limited investor attention to a linked firm's announcement (i.e., customers' unexpected earnings news) leads to market underreactions. This underreaction generates predictable stock returns for suppliers following the customers' earnings announcements. That is, stock prices do not promptly incorporate information from linked firms, generating substantial abnormal returns.

Data and Research Design

Sample Selection

To empirically investigate the relation between the supplier's cumulative abnormal returns surrounding (following) linked customers' earnings announcements and an earnings surprise for the customer, I obtained data from three sources: Compustat segment, which provides linked data for suppliers and their principal customers; CRSP, which provides information on stock returns; and Institutional Brokers' Estimate System (IBES), which provides data on quarterly earnings and the timing of announcements.

According to Regulation SFAS No. 131, firms must periodically release their financial information for any industry segment that comprises more than 10% of consolidated annual sales and for any linked customer that represents more than 10% of total reported sales. Based on the Compustat segment file for each firm, I inspect whether the customer is another company listed in the *CRSP*, Compustat, and IBES files by matching the customer name, and I assign it the corresponding *CRSP* permno number to ensure that customers are matched to the appropriate stock returns and financial information.⁸

I extract stock returns from *CRSP* based on the announcement dates of both suppliers and their linked customers at the same fiscal quarter end. To construct the unexpected earnings, I require the actual earnings and analyst forecasts. The IBES unadjusted individual analyst forecasts for quarterly earnings per share (EPS) are based on the number of shares outstanding on the estimate date. By contrast, the actual reported EPS are based on the number of shares outstanding on the earnings report date. To ensure that both estimated and actual EPS are based on the same number of shares outstanding, I use the *CRSP* cumulative adjustment split factor extracted from the *CRSP* daily stock files. Merging these data with the unadjusted detailed history and the data from the actual files in the IBES database, I generate the final sample of 10,207 firm-quarter observations for the period 1983–2011, which cover a total of 1,083 unique customer-supplier relationships.

Research Design

I examine market reactions to earnings news involving linked customers by estimating several specifications of the following model:

$$CAR[-1,1] = \alpha + \beta_1 CUE + \beta_2 Z + Industry\ Effect + Year\ Effect + \varepsilon, \quad (1)$$

$$CAR[2,61] = \alpha + \beta_1 CUE + \beta_2 Z + Industry\ Effect + Year\ Effect + \varepsilon, \quad (2)$$

where the dependent variable $CAR[-1, 1]$ ⁹ is defined as a supplier's 3-trading-day cumulative abnormal returns around the customer's earnings announcement and $CAR[2, 61]$ as the supplier's subsequent 60-trading-day cumulative abnormal returns after the customer's earnings announcement. The independent variable *CUE* is the customer's unexpected earnings, defined as the actual earnings per share subtracted from the median of the individual analyst forecasts,¹⁰ normalized by the stock price on the date of the fiscal quarter end.

Z is a vector of control variables that are routinely used in return anomaly regressions (e.g., Ramnath, 2002; Cohen and Frazzini, 2008; Hirshleifer, Lim, and Teoh, 2009).¹¹ *ADISTANCE* denotes the absolute value of the reporting lag between the supplier and its linked customers. *FEARLY* is an indicator variable that equals 1 if the supplier announces earnings earlier than its linked customers for the same quarter and that equals 0 otherwise. *CEARLY* is an indicator variable that equals 1 if linked customers announce earnings earlier than the supplier for the same quarter and that equals 0 otherwise. *CBNEWS* is an indicator variable that equals 1 if the customer's unexpected earnings are negative and 0 otherwise. *CANALYSTS* represents the number of analysts following the customer. *PERCRET* represents the ratio of stock returns surrounding the earnings announcement date of linked customers to that of suppliers. *PERCSALE* represents the logarithm of the ratio of sales of economically linked customers to that of suppliers. *PERSIZE* is the logarithm of the ratio of market capitalization of linked customers to that of suppliers, where market capitalization is defined as price times the number of shares outstanding at the end of the fiscal quarter. *MKTRET* represents the market return on the S&P 500 Index surrounding the customers' earnings announcement date.

⁸ Customers for which I could not identify a unique match are excluded from the sample.

⁹ To more accurately assess the potential trading profits of investors, researchers should capture the immediate stock price response to earnings using days -1 , 0 , and $+1$ and the delayed stock price response to earnings beginning returns on day $+2$ (Battalio and Mendenhall, 2011).

¹⁰ Some studies (e.g., Zhang, 2008) have used the latest individual analyst forecast to compute firm's unexpected earnings. I use the latest analyst consensus forecast in my study. However, the inferences are unchanged if I use the latest individual analyst forecast as a proxy for market expectations.

¹¹ See also Givoly and Palmon (1982), Chambers and Penman (1984), Atiase, Bamber, and Tse (1989), Chae (2005), which focus on good (bad) news that is released early (late).

SAMEDAY is an indicator variable that equals 1 if both the supplier and its linked customers report their earnings on the same day and that equals 0 otherwise. SAMEINDUSTRY is an indicator variable that equals 1 if both the supplier and its linked customers are in the same industry and that equals 0 otherwise. I also control for industry (according to the 10-industry classification in Fama and French (1997)) and year effects. ε denotes the error term. Details regarding the construction of the variables are also given in the Appendix.

Table 1 presents the summary statistics for the variables for the full sample. Panel A reports the number of observations, the mean, median, 25th and 75th percentile, and the standard deviation for all variables used in the main analysis. There are 1,083 unique customer-supplier relationships for the period 1983–2011 in the final sample. On average, the supplier’s 3- and 60-trading-day cumulative abnormal returns around and following the linked customer’s earnings announcement are 0.2% and 1.8%, respectively, wherein the customer’s 3- and 60-trading-day cumulative abnormal returns following customer earnings announcements are 0.1% and 0.9%, respectively. This evidence shows that with respect to customer earnings announcements, supplier’s immediate and delayed returns are larger than customer’s immediate and delayed returns. The average unexpected earnings for customers in the sample is –0.1%. The indicator variables of customer bad earnings news (0.359) and supplier bad earnings news (0.783) show that bad earnings news is observed mostly for suppliers in the sample. On average, there are 20 analysts following the customer and 7 analysts following the supplier, respectively. The ratio of the supplier’s abnormal returns surrounding the earnings announcement date to its linked customers is 0.154. The market value of a customer’s equity is about \$20 million, wherein the market value of a supplier is around \$0.6 million (i.e., firm size of customers is 33 times larger than suppliers’ firm size). Linked customer sales in the sample represent 12% of a supplier’s total reported sales on average, which is in line with the reporting requirements by Regulation SFAS No. 131. The reporting lag between suppliers and their linked customers is about 12 days. With regard to early, late, and same-day announcements, 66.2% of sample firms are linked customers who report their quarterly earnings earlier than suppliers, 27.3% of sample firms are suppliers who report their quarterly earnings earlier than their linked customers, and 6.5% are suppliers who release their quarterly earnings on the same day as their linked firms.

Table 1 Summary Statistics

This table presents the summary statistics for all variables used in the main regression analysis of the impact of customer earnings surprises on supplier returns and the sample distribution by the percentage size, sales, and analyst coverage to the customer. The sample consists of 1,083 unique customer-supplier relationships for the period 1983–2011. The appendix outlines the definitions and data sources for the variables. Panel A reports the number of observations, the mean, median, 25th and 75th percentile, and the standard deviation for all variables used in the main analysis. Panel B reports supplier and customer returns in different CUE quintiles. Panels C, D, and E show two portfolios of CUE, supplier and customer returns based on the percentage size, sales, analyst coverage to the customer, respectively.

VARIABLES	N	Mean	SD	P25	Median	P75
<i>Panel A: Summary statistics for full sample</i>						
CAR[-1, 1]	10176	0.002	0.068	-0.027	-0.001	0.028
CAR[2, 61]	10167	0.018	0.281	-0.113	0.014	0.141
CCAR[-1, 1]	10175	0.001	0.064	-0.028	0.000	0.030
CCAR[2, 61]	10172	0.009	0.159	-0.065	0.007	0.084
CUE	10206	-0.001	0.062	0.000	0.000	0.002
CBNEWS	10207	0.359	0.480	0.000	0.000	1.000
CANALYSTS	10207	20.319	15.415	9.000	17.000	28.000
CEARLY	10207	0.662	0.473	0.000	1.000	1.000
CSIZE	10206	16.844	1.723	15.791	17.101	18.117
FBNEWS	10207	0.783	0.412	1.000	1.000	1.000
FANALYSTS	10207	7.237	7.978	2.000	5.000	9.000
FEARLY	10207	0.273	0.445	0.000	0.000	1.000
FSIZE	10146	13.331	1.716	12.147	13.167	14.390

ADISTANCE	10207	11.526	11.029	4.000	8.000	15.000
MKTRET	10207	0.000	0.014	-0.006	0.001	0.007
PERCRET	9791	0.154	11.036	-0.600	0.033	0.977
PERCSALE	10174	-2.120	1.239	-2.809	-2.213	-1.537
PERCSIZE	10145	3.515	2.326	1.740	3.685	5.233
SAMEDAY	10207	0.065	0.246	0.000	0.000	0.000
SAMEINDUSTRY	10207	0.098	0.297	0.000	0.000	0.000

	Mean Percentage	CUE	CAR[-1, 1]	CAR[2, 61]	CCAR[-1, 1]	CCAR[2, 61]
<i>Panel B: Supplier and customer returns by customer earnings surprises</i>						
1		-0.016	0.002	0.026	-0.019	0.019
2		0.000	0.004	0.009	-0.005	0.000
3		0.000	0.003	0.009	0.001	0.004
4		0.001	0.000	0.016	0.006	0.004
5		0.009	0.003	0.028	0.021	0.018
<i>Panel C: Supplier and customer returns by the percentage size to the customer</i>						
1	1.117	-0.002	0.002	0.005	0.001	0.011
2	1.449	0.000	0.003	0.030	0.000	0.007
<i>Panel D: Supplier and customer returns by the percentage sales to the customer</i>						
1	0.056	0.000	0.003	0.015	0.000	0.011
2	1.738	-0.002	0.002	0.020	0.002	0.007
<i>Panel E: Supplier and customer returns by the percentage analysts coverage to the customer</i>						
1	1.575	-0.001	0.002	0.013	0.000	0.010
2	11.176	-0.002	0.002	0.023	0.001	0.008

Panel B reports supplier and customer returns in different customer earnings surprise quintiles. It shows that suppliers in the second quintile of customer earnings surprises have the highest average abnormal return (0.4%) over the customer's 3-day event window, whereas customers in the lowest (highest) quintile of this measure have the lowest (highest) average abnormal return of -1.9% (2.1%) over the customer's 3-day event window. With respect to the delayed returns for both suppliers and customers, the highest average abnormal returns are likely distributed to the lowest and highest quintiles. Panels C through E show two portfolios of customer earnings surprises, supplier returns, and customer returns based on the percentage size, sales, and analyst coverage to the customer, respectively. The evidence shows that the reactions of the supplier's delayed returns are the strongest for firms in the higher percentage of size, sales, and analyst coverage to the customer, suggesting that the higher the percentage size, sales, and analyst coverage to the customer, the more important the customer is to the supplier. Consistent with this interpretation, I later show that firms that are in the portfolio with higher percentage size, sales, and analyst coverage to the customer play an important role in transferring earnings-related information to supplier firms.

Overall, the evidence shows that the average immediate and delayed stock reactions of supplier returns to the customer earnings announcement are stronger than customer's immediate and delayed return reactions to the customer earnings announcement. The evidence suggests that there are observable post-customer earnings announcement supplier returns generated after the customer earnings announcement date.

Empirical Results

Since firm-specific information, such as customer earnings announcements, recurs systematically and induces changes in the valuation of economically related firms, it might have the potential to play a substantial role in the supplier's returns. To gain insight into the relation between

the stock returns of a supplier and the earnings news of its linked customers, I investigate the immediate (delayed) price reaction of the supplier surrounding (following) the customer announcements to the unexpected earnings of its linked customers. If investors are aware of the ex ante customer-supplier links, the stock returns of the supplier will fully adjust when the information about its linked customers is released into the market. If investors do not fully react to information released by the supplier's linked customers, one can expect to observe predictable stock returns during the period following the customer announcements.

Supplier Stock Price Reactions to Earnings News Involving Linked Customers

Table 2 reports the regression results. Columns 1 to 4 present the specifications with the supplier's 3-trading-day cumulative abnormal returns surrounding the earnings announcement date for its linked customers, CAR [-1, 1], as the dependent variable. Each regression employs control variables that have the potential to influence supplier's immediate returns. Column 1 provides the baseline result for the immediate price reaction of a supplier, surrounding the customer announcements, to the unexpected earnings of its customers. I control for CBNEWS, CANALYSTS, PERCSALE, PERCSIZE, SAMEDAY, MKTRET, and their interaction terms with CUE, as well as industry and year effects. The coefficient of CUE is 0.106 and is statistically significant at the 5% level, confirming the presence of the immediate return predictability of a supplier around the customer announcements in my sample. To eliminate the concern that the impact of customer earnings surprises on the supplier's immediate stock returns is driven by the timing and order of announcements, I also control for other correlated factors, such as ADISTANCE, FEARLY, and CEARLY. The coefficient of CUE in column 2 is significantly positive at 0.094, which suggests that suppliers' stock returns are immediately responsive to the earnings announcements of their customers. In column 3, I add a control variable of SAMEINDUSTRY. The coefficient of CUE is significantly positive at 0.093. To mitigate the concern that my results are not driven by stock returns of both suppliers and customers, I also control for the ratio of stock returns of linked customers to suppliers on their own earnings announcement dates respectively (i.e., PERCRET) and its interaction term with CUE. The coefficient of CUE in column 4 is significantly positive at 0.120, which confirms the immediate responsiveness of stock returns to unexpected customers' earnings.

Table 2 Test of Supplier's Stock Price Reaction Surrounding the Customer Announcements

This table presents the estimation results obtained by regressing the supplier's cumulative abnormal returns around the customer's earnings announcement (CAR [-1, 1]) on the customer's unexpected earnings (CUE), on several controls, and on the interaction terms between CUE and controls. The appendix outlines the definitions and data sources for the regression variables. For the sake of brevity, industry and year controls are included, but the results are not tabulated; the industry controls are based on Fama and French's (1997) 10-industry classification. *t*-statistics based on robust standard errors adjusted for clustering by firms are reported in parentheses. ***, **, *, indicate statistical significant at the 0.01, 0.05 and 0.10 level, respectively.

$H_{1a}: CAR[-1,1] = \alpha + \beta_1 CUE + \beta_2 Z + Industry\ Effect + Year\ Effect + \varepsilon$				
VARIABLES	(1) CAR[-1, 1]	(2) CAR[-1, 1]	(3) CAR[-1, 1]	(4) CAR[-1, 1]
CUE	0.106** (2.30)	0.094** (2.03)	0.093* (1.94)	0.120** (2.07)
PERCRET				0.000 (0.28)
CUE × PERCRET				-0.004 (-0.45)
SAMEINDUSTRY			0.001 (0.53)	0.002 (0.67)
CUE × SAMEINDUSTRY			-0.052 (-0.57)	-0.081 (-0.91)

CUE × ADISTANCE × FEARLY		0.002 (0.95)	0.002 (0.80)	0.001 (0.38)
CUE × ADISTANCE × CEARLY		0.003 (1.01)	0.003 (1.00)	0.003 (1.08)
CBNEWS	0.001 (0.71)	0.001 (0.73)	0.001 (0.71)	0.001 (0.73)
CANALYSTS	0.000 (0.75)	0.000 (0.77)	0.000 (0.78)	0.000 (0.51)
PERCSALE	-0.000 (-0.76)	-0.000 (-0.77)	-0.000 (-0.81)	-0.000 (-0.78)
PERCSIZE	0.000 (0.83)	0.000 (0.83)	0.000 (0.91)	0.000 (0.94)
SAMEDAY	0.001 (0.24)	0.001 (0.23)	0.001 (0.20)	-0.001 (-0.18)
MKTRET	0.140* (1.89)	0.140* (1.88)	0.140* (1.88)	0.159** (2.09)
CUE × CBNEWS	-0.068 (-1.37)	-0.077 (-1.54)	-0.077 (-1.56)	-0.102 (-1.48)
CUE × CANALYSTS	-0.000 (-0.00)	0.000 (0.03)	0.000 (0.04)	-0.000 (-0.12)
CUE × PERCSALE	0.029** (2.47)	0.035** (2.33)	0.033* (1.80)	0.035* (1.67)
CUE × PERCSIZE	-0.003 (-0.26)	-0.001 (-0.07)	-0.002 (-0.16)	0.010 (0.76)
CUE × SAMEDAY	-0.168* (-1.89)	-0.161* (-1.77)	-0.151 (-1.52)	-0.150 (-1.53)
CUE × MKTRET	0.880 (0.60)	0.531 (0.36)	0.523 (0.36)	-0.349 (-0.24)
Intercept	-0.007 (-0.72)	-0.005 (-0.47)	-0.005 (-0.52)	-0.006 (-0.35)
N	10,110	10,110	10,110	9,735
R ²	0.008	0.008	0.008	0.009

Remarkably, the coefficients for the interaction term (CUE × PERCSALE) are positive at the 10% level in columns 1 through 4. The total effects of CUE on CAR[-1,1], even when there is higher percentage sales to customer, are still positive with the magnitude of 0.135, 0.129, 0.126, and 0.155 in columns 1 through 4, respectively. These results indicate that the higher the percentage sales to customer, the stronger the effect of customer earnings surprises on supplier's immediate returns.

Taken together, the evidence in Table 2 indicates that the cumulative abnormal return of a supplier around the customer announcements is positively related to the unexpected earnings of linked customers.

Supplier Stock Price Reactions Following the Customer Announcements

The immediate responsiveness of stock returns surrounding the customer announcements, as indicated above, offers important evidence of the market reactions to earnings news involving the customers of suppliers. However, Cohen and Frazzini (2008) report that stock prices do not fully

reflect news involving related firms, which generates predictable subsequent price moves. Accordingly, I further investigate cumulative stock return responses of a supplier to customers' unexpected earnings following the customer announcements.

Table 3 reports the estimation results. Columns 1 to 4 present the specifications with the abnormal returns of a supplier cumulated 60-trading-day after the earnings announcement date for its linked customers, CAR [2, 61], as the dependent variable. The independent and control variables used in the delayed stock return analysis are the same as in the immediate stock return analysis. The coefficients of CUE in columns 1 to 4 are all positive and significant at the 1% level with magnitude of 0.724, 0.843, 0.834, and 0.849, respectively. The positive and statistically significant coefficient is in line with the hypothesis, and suggests that the stock reactions of a supplier after linked customer earnings announcements to unexpected customers' earnings are substantial.

Table 3 Test of Supplier's Stock Price Reaction Following the Customer Announcements

This table presents the estimation results obtained by regressing the supplier's cumulative abnormal returns following the customer's earnings announcement (CAR [2, 61]) on the customer's unexpected earnings (CUE), on several controls, and on the interaction terms between CUE and controls. Appendix outlines the definitions and data sources for the regression variables. For the sake of brevity, the industry and year controls are included but are not tabulated; the industry controls are based on Fama and French's (1997) 10-industry classification. *t*-statistics based on robust standard errors adjusted for clustering by firms are reported in parentheses. ***, **, *, indicate statistical significant at the 0.01, 0.05 and 0.10 level, respectively.

$H_{1b}: CAR[2,61] = \alpha + \beta_1 CUE + \beta_2 Z + Industry\ Effect + Year\ Effect + \varepsilon$				
VARIABLES	(1) CAR[2, 61]	(2) CAR[2, 61]	(3) CAR[2, 61]	(4) CAR[2, 61]
CUE	0.724*** (3.50)	0.843*** (4.32)	0.834*** (4.27)	0.849*** (3.16)
PERCRET				0.000 (1.17)
CUE × PERCRET				0.016 (0.44)
SAMEINDUSTRY			0.003 (0.32)	-0.000 (-0.01)
CUE × SAMEINDUSTRY			-0.235 (-0.56)	-0.245 (-0.56)
CUE × ADISTANCE × FEARLY		-0.025* (-1.74)	-0.026* (-1.76)	-0.023 (-1.32)
CUE × ADISTANCE × CEARLY		-0.020 (-1.45)	-0.020 (-1.45)	-0.014 (-1.11)
CBNEWS	0.007 (1.09)	0.007 (1.11)	0.007 (1.08)	0.010 (1.52)
CANALYSTS	-0.000* (-1.79)	-0.000* (-1.82)	-0.000* (-1.81)	-0.000* (-1.85)
PERCSALE	0.004 (1.48)	0.004 (1.48)	0.004 (1.45)	0.004 (1.61)
PERCSIZE	0.007*** (5.00)	0.007*** (5.01)	0.008*** (4.98)	0.008*** (4.83)
SAMEDAY	-0.019**	-0.019*	-0.019**	-0.022**

	(-2.00)	(-1.96)	(-1.98)	(-2.17)
MKTRET	0.360	0.364	0.364	0.296
	(1.28)	(1.30)	(1.29)	(1.04)
CUE × CBNEWS	-0.600**	-0.509**	-0.508**	-0.552*
	(-2.25)	(-2.27)	(-2.27)	(-1.80)
CUE × CANALYSTS	0.004	0.004	0.004	0.004
	(0.72)	(0.69)	(0.70)	(0.57)
CUE × PERCSALE	0.246***	0.173***	0.161***	0.150**
	(3.49)	(3.40)	(2.72)	(2.33)
CUE × PERCSIZE	0.045	0.014	0.010	0.023
	(1.13)	(0.32)	(0.22)	(0.49)
CUE × SAMEDAY	-0.899***	-0.943***	-0.900***	-0.849***
	(-3.09)	(-3.82)	(-3.35)	(-3.13)
CUE × MKTRET	2.419	6.220	6.190	5.400
	(0.48)	(1.19)	(1.19)	(1.00)
Intercept	0.033	0.036	0.036	0.010
	(0.71)	(0.79)	(0.79)	(0.19)
N	10,101	10,101	10,101	9,727
R ²	0.030	0.031	0.031	0.031

Likewise, the coefficients for the interaction term (CUE × PERCSALE) are positive at the 5% level in columns 1 through 4. The total effects of CUE on CAR[2,61], even when there is higher percentage sales to customer, are still positive with the magnitude of 0.970, 1.016, 0.995, and 0.999 in columns 1 through 4, respectively. Similar as the effect of customer earnings surprises on supplier's immediate returns in Table 2, the higher the percentage sales to customer even stronger the effect of customer earnings surprises on supplier's delayed returns. The evidence indicates that the presence of large percentage sales to customer is associated with stronger supplier's immediate and delayed price reactions to customer earnings surprises.

Interestingly, the coefficients of the interaction term (CUE × CBNEWS) in Table 2 are all insignificant, however, the coefficients of the interaction term (CUE × CBNEWS) in Table 3 are all negative and significant at the 10% level. The sum of coefficients on CUE and (CUE × CBNEWS) are still positive with the magnitude of 0.124, 0.334, 0.326, and 0.297 in columns 1 to 4 in Table 3, respectively. Since the measure of CBNEWS is a binary variable for when customer earnings news is negative, the coefficients on CUE in Tables 2 and 3 should reflect the reaction of supplier returns to positive customer earnings surprises (good news) only, and the sum of the coefficients on CUE and the interaction term (CUE × CBNEWS) will reflect the reaction of supplier returns to negative customer earnings surprises (bad news). Thus, supplier's immediate reaction to good news is positive and significant, but there is no difference in supplier's delayed reaction to CUE based on the sign of the news. This evidence shows that supplier returns have a stronger, if any, delayed reaction to negative customer earnings news, which is in line with the notion of slow diffusion of bad news (Hong and Stein, 1999; Hong et al., 2000; Hou, 2007).

Remarkably, the coefficients on the interaction term (CUE × SAMEDAY) are negative at the 1% level and approach the magnitude of the coefficients on CUE in columns 1 through 4, respectively. Since the sum of the coefficients on CUE and (CUE × SAMEDAY) reflects the reaction of supplier returns to CUE when customers and suppliers report earnings on the same day, the evidence shows that supplier returns have weaker delayed reaction to customer earnings news. I interpret this as evidence of the investor distraction hypothesis (Hirshleifer, et al., 2009), whereby customer earnings news draws investor attention away from supplier when customers and suppliers report earnings on the same day, and further generates less supplier returns.

Overall, the results in Tables 2 and 3 show that stock prices do not promptly incorporate information from linked firms, which in turn generates the abnormal returns for suppliers cumulated after earnings announcements by related customers. These results are consistent with the notion that systematic limited attention to a given piece of information predicts return forecastability with respect to the impact on firm value of the piece of information being ignored, and thus provide support to the hypothesis that limited investor attention to economic links causes market underreactions.

Enhancing Effect of Customer Earnings Surprises on the Customer-Supplier Returns

The limited investor attention hypothesis by Cohen and Frazzini (2008) argues that customer returns predict future supplier's returns because investors pay limited attention to the customer-supplier link. Though the second hypothesis discussed above supports a delayed supplier reaction to customer earnings announcements, it is of great interest to identify what role, if any, customer earnings surprises play in the customer-supplier return predictability.

Under the scheme that customer earnings surprises leads to customer returns and the latter, in turn, affects supplier returns, customer earnings surprises might have the potential to play a substantial role in the relationship between customer and supplier returns. Along this view, the predictability of supplier returns is simply the consequence of increases in customer returns due to customer earnings surprises. Thus, I argue that investor's underreaction to the customer's earnings surprise leads to a drift in the customer's post-earnings announcement returns (Bernard and Thomas, 1989), which in turn is reflected in the supplier's returns due to investors not taking economic links into account rationally and efficiently (Cohen and Frazzini, 2008).

In attempt to investigate the impact of customer earnings surprises on customer-supplier returns, I estimate a regression including the cumulated post-earnings announcement return of the customer (CCAR [2, 61]) and its interaction term with the customer's standardized unexpected earnings (SUE).¹² SUE is defined as decile-adjusted unexpected earnings and is obtained by ranking CUE into ten deciles based on the sample distribution of customer earnings surprises by quarter in each industry and then scaling them between 0 and 1 (Bernard and Thomas, 1990; Bartov et al., 2000; Zhang, 2008). I use the interaction term between customer returns and the standardized unexpected earnings (CCAR [2, 61] \times SUE) to test the role of customer earnings surprises in the relationship between customer returns and supplier returns. In line with the second hypothesis, I expect that the coefficient of the interaction term to be positive, suggesting that market underreaction to earnings surprises enhances the positive effect of customer returns on supplier returns due to limited attention to customer-supplier economic link.

$$CAR[2,61] = \alpha + \beta_1 CUE + \theta_1 CCAR[2,61] + \theta_2 (CCAR[2,61] \times SUE) + \beta_2 W + Industry\ Effect + Year\ Effect + \varepsilon \quad (3)$$

The results in Table 4 show that the coefficients for CUE are positive and significant at the 5% level whereas for the interaction terms (CCAR [2, 61] \times SUE) they are also positive and significant at the 5% level. Remarkably, the total effects of customer earnings surprises on supplier returns, even when there is customer returns, are still positive and significant at the 1% and are enhanced to 0.843, 0.970, 0.968, and 0.814 in columns 1 to 4, respectively. It is remarkable that the direct effects of customer returns on supplier returns are positive at the 1% level, in which the customer's post-earnings drift is reflected significantly in the supplier's returns. This finding is consistent with the limited investor attention to economic links hypothesis (Cohen and Frazzini, 2008). Overall, these results provide support for the positive enhancing effect of customer earnings surprises in the relationship between customer returns and supplier returns, implying that the customer's earnings surprise is a predominant source of the cross-firm return predictability.

Table 4 Enhancing Effect of Customer Earnings Surprises on the Customer-supplier Returns

This table presents the estimation results for the enhancing effect of customer earnings surprises on the relationship between customer returns (CCAR [2, 61]) and supplier returns (CAR [2, 61]). The models augment the basic specifications of H_{1b} in

¹² Using the customer standardized unexpected earnings (SUE) instead of customer unexpected earnings (CUE), I focus on the enhancing role when the degree of customer earnings surprises is high enough.

Table 3 by including the cumulated post-earnings announcement return of the customer (CCAR [2, 61]) and its interaction term with the customer's standardized unexpected earnings (SUE). Appendix outlines the definitions and data sources for the regression variables. For the sake of brevity, industry and year controls are included but are not tabulated, and the industry controls are based on Fama and French's (1997) 10-industry classification. *t*-statistics based on robust standard errors adjusted for clustering by firms are reported in parentheses. ***, **, *, indicate statistical significant at the 0.01, 0.05 and 0.10 level, respectively.

VARIABLES	(1) CAR[2, 61]	(2) CAR[2, 61]	(3) CAR[2, 61]	(4) CAR[2, 61]
CUE	0.680*** (2.93)	0.815*** (3.53)	0.813*** (3.49)	0.662** (2.51)
CCAR[2, 61]	0.202*** (4.90)	0.208*** (5.00)	0.207*** (5.01)	0.208*** (5.01)
CCAR[2, 61] × SUE	0.163** (2.37)	0.155** (2.23)	0.155** (2.24)	0.152** (2.18)
PERCRET				0.000 (1.10)
CUE × PERCRET				0.021 (0.57)
SAMEINDUSTRY			0.002 (0.26)	-0.001 (-0.09)
CUE × SAMEINDUSTRY			-0.069 (-0.18)	-0.115 (-0.30)
CUE × ADISTANCE × FEARLY		-0.028* (-1.89)	-0.028* (-1.87)	-0.027 (-1.49)
CUE × ADISTANCE × CEARLY		-0.023* (-1.70)	-0.023* (-1.70)	-0.018 (-1.42)
CBNEWS	0.008 (1.25)	0.008 (1.26)	0.008 (1.25)	0.010 (1.60)
CANALYSTS	-0.000* (-1.91)	-0.000* (-1.95)	-0.000* (-1.94)	-0.000* (-1.94)
PERCSALE	0.004 (1.53)	0.004 (1.53)	0.004 (1.51)	0.004* (1.68)
PERCSIZE	0.008*** (5.30)	0.008*** (5.32)	0.008*** (5.26)	0.008*** (5.03)
SAMEDAY	-0.019** (-1.98)	-0.018* (-1.94)	-0.018* (-1.94)	-0.021** (-2.13)
MKTRET	0.344 (1.25)	0.348 (1.27)	0.348 (1.27)	0.277 (1.00)
CUE × CBNEWS	-0.424* (-1.68)	-0.316 (-1.39)	-0.317 (-1.39)	-0.147 (-0.49)
CUE × CANALYSTS	0.005 (0.78)	0.005 (0.76)	0.005 (0.76)	0.004 (0.57)
CUE × PERCSALE	0.237*** (3.03)	0.153*** (2.96)	0.150** (2.40)	0.111* (1.76)
CUE × PERCSIZE	0.007	-0.030	-0.031	-0.046

	(0.15)	(-0.65)	(-0.64)	(-0.98)
CUE × SAMEDAY	-1.029***	-1.083***	-1.069***	-1.012***
	(-3.39)	(-4.62)	(-4.07)	(-3.77)
CUE × MKTRET	4.844	9.227**	9.215**	11.161**
	(1.01)	(1.99)	(1.99)	(2.17)
Intercept	0.012	0.015	0.015	-0.061
	(0.22)	(0.29)	(0.29)	(-1.24)
N	10,098	10,098	10,098	9,724
R ²	0.055	0.055	0.055	0.056

Robustness Checks

The focus of H_{1a} and H_{1b} is on whether investors have limited attention with respect to the information contained in customer earnings announcements for suppliers when the ex ante customer-supplier links between firms are publicly available.

To address the possible effect of different horizons on return sensitivities, I compute supplier stock returns following the customer announcements over 30-, 45-, 60-, 75-, and 90-trading-day horizons. I find that in using these contrasting horizons, the results are quite similar. I therefore conclude that the delayed returns over different horizons are not likely to affect the results.

To assess how my previous inferences change depending on how important the customer is to the supplier, I repeat the main analyses using alternative sample selections based on the percentage size, sales, and analyst coverage to the customer. The subsamples are based on the percentage size to the customer, the percentage sales to the customer, and the percentage analyst coverage to the customer. These results are consistent with evidence in Table 3 and further address the importance of the percentage size, sales, and analyst coverage to the customer on determining the predictive power of customer earnings surprises on supplier returns.

Overall, these results¹³ provide evidence that customer earnings announcements convey information for suppliers depending on how important the customer is to the supplier, and thus provide further support for investor inattention to customer-supplier information transfers.

Concluding Remarks

The limited investor attention hypothesis proposes that limited investor attention to firm-specific information of economically linked firms (Cohen and Frazzini, 2008), a large number of same-day earnings announcements made by other firms (Hirshleifer, Lim, and Teoh, 2009), Friday earnings announcements (Dellavigna and Pollet, 2009), and stock recommendations (Loh, 2010) cause market underreactions to new information, generating predictable returns across firms.

This study investigates whether limited investor attention to linked customers' earnings announcements leads to market underreactions. I show that the cumulative abnormal stock returns of a supplier surrounding and following its linked customers' earnings announcement date are positively related to customers' earnings news, which suggests that investors are inattentive to customer-supplier information transfers from earnings announcements. The evidence consistently supports the limited attention hypothesis. Investors pay limited attention to economic links and are slow to update their expectations regarding future earnings upon receiving new information from economically related firms. When customer news is released into the market, investors pay limited attention to the ex ante economic link. Consequently, the price of the supplier's stock does not promptly adjust; instead, the more substantial reaction is predictable following the customers' surprising earnings.

Under the scheme that customer earnings surprises leads to customer returns (e.g. Bernard and Thomas, 1989) and the latter, in turn, generates supplier returns (Cohen and Frazzini, 2008), I

¹³ These results are not included for the sake of brevity and are available upon request.

find that customer earnings surprises play a substantial role in the relationship between customer returns and supplier returns. This enhancing effect of customer earnings surprises on the customer-supplier returns provides strong support that the customer's earnings surprise is a predominant source of the cross-firm return predictability.

Appendix Description of the Variables

Variables	Description and Source
<i>Panel A: Cumulative abnormal returns</i>	
CAR[-1,1]	Supplier's 3-trading-day cumulative abnormal returns around the customer's earnings announcement. <i>Source: author's calculations based on Compustat segment, CRSP, and IBES.</i>
CAR[2,61]	Supplier's subsequent 60-trading-day cumulative abnormal returns after the customer's earnings announcement. <i>Source: author's calculations based on Compustat segment, CRSP, and IBES.</i>
CCAR[-1, 1]	Customer's 3-trading-day cumulative abnormal returns surrounding the customer's earnings announcement. <i>Source: author's calculations based on CRSP and IBES.</i>
CCAR[2, 61]	Customer's subsequent 60-trading-day cumulative abnormal returns after the customer's earnings announcement. <i>Source: author's calculations based on CRSP and IBES.</i>
<i>Panel B: Unexpected earnings</i>	
CUE	Customer's unexpected earnings is defined as the actual earnings per share subtracted from the median of individual analyst forecasts scaled by the end of the quarter share price. <i>Source: author's calculations based on IBES and CRSP.</i>
SUE	Customer's standardized unexpected earnings is defined as decile-adjusted unexpected earnings and is obtained by ranking CUE into ten deciles based on the sample distribution of CUE by quarter in each industry and then scaling them between 0 and 1. <i>Source: author's calculations based on IBES and CRSP.</i>
<i>Panel C: Other control variables</i>	
CBNEWS	Indicator variable that equals 1 if the customer's unexpected earnings are negative and 0 otherwise. <i>Source: author's calculations based on IBES.</i>
CANALYSTS	The number of analysts following the customer. <i>Source: author's calculations based on IBES.</i>
CEARLY	Indicator variable that equals 1 if linked customers announce earnings earlier than the supplier for the same quarter and 0 otherwise. <i>Source: author's calculations based on IBES.</i>
CSIZE	Customer's log of market capitalization which is price times the number of shares outstanding at the end of the fiscal quarter. <i>Source: author's calculations based on CRSP.</i>
FBNEWS	Indicator variable that equals 1 if the supplier's unexpected earnings are negative and 0 otherwise. <i>Source: author's calculations based on IBES.</i>
FANALYSTS	The number of analysts following the supplier. <i>Source: author's calculations based on IBES.</i>
FEARLY	Indicator variable that equals 1 if the supplier announces earnings earlier than its linked customers for the same quarter and 0 otherwise. <i>Source: author's calculations based on IBES.</i>
FSIZE	Supplier's log of market capitalization which is price times the number of shares outstanding at the end of the fiscal quarter. <i>Source: author's calculations based on CRSP.</i>
ADISTANCE	The absolute value of the reporting lag between the supplier and its linked customers in each quarter. <i>Source: author's calculations based on Compustat segment, CRSP, and IBES.</i>
MKTRET	Market returns on the S&P 500 Index. <i>Source: CRSP.</i>
PERCRET	Ratio of stock returns of linked customers to suppliers on their earnings announcement date respectively. <i>Source: author's calculations based on IBES and CRSP.</i>
PERCSALE	Log of the ratio of sales of economically linked customers to suppliers. <i>Source: author's calculations based on Compustat segment.</i>
PERCSIZE	Log of the ratio of market capitalization of linked customers to suppliers. <i>Source: author's calculations based on CRSP.</i>
SAMEDAY	Indicator variable that equals 1 if both the supplier and its customers release their earnings on the same day and that equals 0 otherwise. <i>Source: author's calculations based on IBES.</i>
SAMEINDUSTRY	Indicator variable that equals 1 if both the supplier and its linked customers are in the same industry and that equals 0 otherwise. <i>Source: author's calculations based on Compustat segment and Fama and French's (1997) 10-industry classification.</i>

References

- Abarbanell, J. and V. Bernard (1992), 'Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior', *Journal of Finance*, Vol. 47, pp. 1191–1207.
- Atiase, R. K., L. S. Bamber and S. Tse (1989), 'Timeliness of Financial Reporting, the Firm Size Effect, and Stock Price Reactions to Annual Earnings Announcements', *Contemporary Accounting Research*, Vol. 5, pp. 526–552.
- Bae, K. H., A. Ozoguz, H. Tan and T. S. Wirjanto (2012), 'Do Foreigners Facilitate Information Transmission in Emerging Markets', *Journal of Financial Economics*, Vol. 105, pp. 209–227.
- Bae, K. H. and W. Wang (2012), 'What is in a "China" Name? A Test of Investor Attention Hypothesis', *Financial Management*, Vol. 41, pp. 429–455.
- Barber, B. M. and T. Odean (2008), 'All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors', *Review of Financial Studies*, Vol. 21, pp. 785–818.
- Bartov, E., S. Radhakrishnan and I. Krinsky (2000), 'Investor Sophistication and Patterns in Stock Returns after Earnings Announcements', *Accounting Review*, Vol. 75, pp. 43–63.
- Battalio, R. H. and R. R. Mendenhall (2011), 'Post-earnings Announcement Drift: Bounds on Profitability for the Marginal Investor', *Financial Review*, Vol. 46, pp. 513–539.
- Bernard, L. V. and K. J. Thomas (1989), 'Post-earnings-announcement Drift: Delayed Price Response or Risk Premium', *Journal of Accounting Research*, Vol. 27, pp. 1–36.
- , ——— (1990), 'Evidence that Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings', *Journal of Accounting and Economics*, Vol. 13, pp. 305–340.
- Cai, J., M. H. Song and R. A. Walkling (2011), 'Anticipation, Acquisitions, and Bidder Returns: Industry Shocks and the Transfer of Information across Rivals', *Review of Financial Studies*, Vol. 24, pp. 2242–2285.
- Chae, J. (2005), 'Trading Volume, Information Asymmetry, and Timing Information', *Journal of Finance*, Vol. 60, pp. 413–442.
- Chambers, A. E. and S. H. Penman (1984), 'Timeliness of Reporting and the Stock Price Reaction to Earnings Announcements', *Journal of Accounting Research*, Vol. 22, pp. 21–47.
- Cohen, L. and A. Frazzini (2008), 'Economic links and predictable returns', *Journal of Finance*, Vol. 63, pp. 1977–2011.
- Da, Z., J. Engelberg and P. Gao (2011), 'In search of Attention', *Journal of Finance*, Vol. 66, pp. 1461–1500.
- Dellavigna, S. and J. M. Pollet (2007), 'Demographics and Industry Returns', *American Economics Review*, Vol. 97, pp. 1167–1702.
- , ——— (2009), 'Investor Inattention and Friday Earnings Announcements', *Journal of Finance*, Vol. 64, pp. 709–749.
- Desir, R. (2013), 'How Do Managers of Non-Announcing Firms Respond to Intra-Industry Information Transfers', *Journal of Business Finance & Accounting*, Vol. 39, pp. 1180–1213.
- Fama, E. F. and K. R. French (1997), 'Industry Costs of Equity', *Journal of Financial Economics*, Vol. 43, pp. 153–193.
- Freeman, R. N. and S. Y. Tse (1992), 'An Earnings Prediction Approach to Examining Intercompany Information Transfers', *Journal of Accounting and Economics*, Vol. 15, pp. 509–523.
- Gilbert, T., S. Kogan, L. Lochstoer and A. Ozyildirim (2012), 'Investor Inattention and the Market Impact of Summary Statistics', *Management Science*, Vol. 58, pp. 336–350.
- Givoly, D. and D. Palmon (1982), 'Timeliness of Annual Earnings Announcements: Some Empirical Evidence', *Accounting Review*, Vol. 57, pp. 486–508.
- Hirshleifer, D. and S. H. Teoh (2003), 'Limited Attention, Information Disclosure, and Financial Reporting', *Journal of Accounting and Economics*, Vol. 36, pp. 337–386.
- , S. Lim and ——— (2009), 'Driven to Distraction: Extraneous Events and Underreaction to Earnings News', *Journal of Finance*, Vol. 64, pp. 2289–2325.
- Hong, H. and J. C. Stein (1999), 'A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets', *Journal of Finance*, Vol. 54, pp. 2143–2184.

- , T. Lim and ——— (2000), ‘Bad News Travels Slowly: Size, Analyst coverage, and the Profitability of Momentum Strategies’, *Journal of Finance*, Vol. 55, pp. 265–295.
- , W. Torous and R. Valkanov (2007), ‘Do Industries Lead the Stock Market’, *Journal of Financial Economics*, Vol. 83, pp. 367–396.
- Hou, K. (2007), ‘Industry Information Diffusion and the Lead-lag Effect in Stock Returns’, *Review of Financial Studies*, Vol. 20, pp. 1113–1138.
- , L. Peng and W. Xiong (2009), ‘A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum’, Working Paper (Ohio State University).
- Huberman, G. and T. Regev (2001), ‘Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar’, *Journal of Finance*, Vol. 56, pp. 387–396.
- Jorion, P. and G. Zhang (2010), ‘Information Transfer Effects of Bond Rating Downgrades’, *Financial Review*, Vol. 45, pp. 683–706.
- Kovacs, T. (2011), ‘Intra-industry Information Transfers and the Post-earnings Announcement Drift’, Working Paper (Northeastern University).
- Koch, A. S. and A. X. Sun (2004), ‘Dividend Changes and Persistence of Past Earnings Changes’, *Journal of Finance*, Vol. 59, pp. 2093–2116.
- Loh, R. (2010), ‘Investor Inattention and the Underreaction to Stock Recommendations’, *Financial Management*, Vol. 39, pp. 1223–1251.
- Menzly, L. and O. Ozbas (2010), ‘Market Segmentation and Cross-predictability of Returns’, *Journal of Finance*, Vol. 65, pp. 1555–1580.
- Merton, R. (1987), ‘A Simple Model of Capital Market Equilibrium with Incomplete Information’, *Journal of Finance*, Vol. 42, pp. 483–510.
- Peng, L. (2005), ‘Learning with Information Capacity Constraints’, *Journal of Financial and Quantitative Analysis*, Vol. 40, pp. 307–329.
- , W. Xiong (2006), ‘Investor Attention, Overconfidence and Category Learning’, *Journal of Financial Economics*, Vol. 80, pp. 563–602.
- Prokopczuk, M. (2010), ‘Intra-industry Contagion Effects of Earnings Surprises in the Banking Sector’, *Applied Financial Economics*, Vol. 20, pp. 1601–1613.
- Ramnath, S. (2002), ‘Investor and Analyst Reactions to Earnings Announcements of Related Firms: An Empirical Analysis’, *Journal of Accounting Research*, Vol. 40, pp. 1351–1376.
- Yuan, Y. (2012), ‘Market-wide Attention, Trading, and Stock Returns’, Working Paper (University of Pennsylvania).
- Zhang, Y. (2008), ‘Analyst Responsiveness and the Post-earnings-announcement Drift’, *Journal of Accounting and Economics*, Vol. 46, pp. 201–215.

Modelling the Effect of Transactions Costs on Options-Related Trading Frequency

Abstract

In this paper we examine the relationship between transactions costs and options-related trading frequency. We develop a model which enables us to examine option hedging behaviour through any set of asset price and time paths, including large price changes in small time intervals, within a Black-Scholes numerical solution set. This approach allows for a type of completeness that is not possible with a binomial tree or Monte Carlo simulation. Our model suggests a linear relationship between transactions costs and trading frequency.

1. Introduction

Under the assumptions of the original Black-Scholes (B-S 1973) model for options pricing, there are no transactions costs and trading is continuous in time. When transactions costs are positive, trading at discrete but frequent intervals is prohibitively expensive. The investor who hedges a written position in an option by trading the option and the underlying asset will be forced to trade less frequently, and will face hedging losses as a result. We explore the relation between trading frequency and transactions costs.

Section 2 gives a review of relevant literature. Section 3 analyzes the impact of transactions costs on the B-S model. In Section 4 we provide an overview of the interaction between transactions costs and risk taking. Section 5 discusses the setup of the model, and Section 6 gives an explanation of the model with an outline of some results. We follow in Section 7 with a quantitative analysis of these results. In Section 8 we conclude and suggest possibilities for further research.

2. Literature Review

In the original B-S paper, several assumptions were made that allowed them to solve an option pricing equation without knowledge of risk preferences. Of interest are the assumptions of continuous trading and zero transactions costs. There have been several papers which derive techniques for option pricing when trading is not continuous and transactions costs are not zero.

Leland (1985) incorporates transactions costs into the pricing equation through an adjustment to the volatility term in the asset price process. A more volatile underlying asset requires a larger adjustment in the hedging position, and therefore a higher transactions cost. Our analysis addresses Leland's statement that "transactions costs associated with replicating strategies are path-dependent: they depend not only on the initial and final stock prices, but on the entire sequence of stock prices in between" (Leland, p. 1284). Our model enables traversing through every possible path, and uses the information from all the paths.

Boyle and Vorst (BV) (1992) derived a model including transactions costs in discrete time. BV use the Cox et al. (CRR) (1979) binomial tree model as a basis for their conclusions. The CRR binomial tree

model allows for an option of two price steps at each time step. Our model extends this to allow the price of the asset to be any of a wide number of values at the next time step.

3. Transactions Costs and the Black-Scholes Model

As with the B-S model for options pricing, we start with a portfolio composed of a long position in one unit of a call option, and short position in some quantity of the asset underlying that option. The value Π of this portfolio can be written

$$\Pi = V(S, t) - \Delta S$$

where V is the option price, S is the underlying asset price and Δ is the quantity of this asset held in a short position. We assume the asset price S to follow the standard Brownian-motion process.

$$(1) \quad dS = \mu S dt + \sigma S z \sqrt{dt}$$

where μ is the drift rate of the asset price and σ is its standard deviation. The term z is a standard normal random variable. Deriving the time path of Π requires the use of stochastic calculus (Itô's lemma). The result is

$$(2) \quad \begin{aligned} d\Pi &= dV - \Delta dS \\ &= V_t dt + V_S dS + \frac{1}{2} \sigma^2 S^2 V_{SS} dt - \Delta dS \end{aligned}$$

The B-S uses the fact, from (2), that if Δ is chosen such that $\Delta = V_S$ then the stochastic term will be eliminated. The portfolio must then earn the riskless rate of return. The resulting equality $d\Pi = r\Pi dt$ gives the Black-Scholes partial differential equation. This derivation assumes no transactions costs and continuous trading so that the investor can maintain $\Delta = V_S$ at all times. When transactions costs are positive, continuous trading becomes infinitely expensive and trades can only be made at discrete intervals. In the time that elapses between these trades, the hedging position Δ is fixed but V_S will change because the underlying asset price changes. This results in a discrepancy between the two values, called a hedging error, denoted ε and defined as $\varepsilon = V_S - \Delta$. This error results in a nonzero coefficient on the stochastic term dS and is the source of risk in the portfolio:

$$(3) \quad d\Pi = V_t dt + \frac{1}{2} \sigma^2 S^2 V_{SS} dt + \varepsilon dS$$

4. Theoretical Basis for Model

First, notice from (1) and (3) that the variance of the portfolio (the coefficient of the stochastic term $z\sqrt{dt}$) is dependent on the magnitude of the product of hedging error ε and asset price S . The absolute value of this product, denoted η , will be called portfolio risk. In our model, investors allow the portfolio risk η to grow to a critical size before eliminating it. This interval of acceptable risk is known as a band of inaction. We assume no other fixed transactions costs and no limits on the divisibility of securities, market features that might otherwise influence an investor's band of inaction. The edge of the band is determined as the

point where the risk has become excessive, and so the transactions cost is paid and the hedging error ε (and thus the risk η) is eliminated. Excessive risk is defined as that for which

$$(4) \quad f(\eta, \gamma) \geq k\eta$$

where γ is a constant of risk aversion, and f is some function describing the investor dollar disutility of risk. The term k is the proportional transaction cost and the right-hand side defines the total transaction cost of eliminating the hedging error ε .

It is worth noting that this f is somewhat related to the Arrow-Pratt coefficient of risk-aversion (Pratt 1964),

$$A(w) = \frac{U''(w)}{U'(w)}$$

where U is a utility function of wealth. In this model, risk aversion is derived from the diminishing marginal returns to wealth. Our function f is also analogous to the component of the mean-variance analysis of a financial asset that describes the negative effect of an increase in volatility on that assets value. If $H(\mu, \sigma^2)$ is a preference function describing the value of an asset, based on its expected return μ and variance σ^2 , our function f would be related to $\partial H / \partial(\sigma^2)$, although with the opposite sign.

The fact that the threshold η^* exists, that there is some amount of risk that is tolerable in a portfolio, but that after a certain point (4) the investor is better-off eliminating it, provides information about the nature of risk attitudes among investors. In order for η^* to exist given the characteristics of the cost formula $k\eta$ on the right-hand side of (4), the function f must have the following properties, properties which allow for the intersection of f and $k\eta$ at some point η^* , and insure that on the interval $(0, \eta^*)$, we have $f < k\eta$.

$$\frac{\partial f}{\partial \eta} \geq 0, \quad \frac{\partial^2 f}{\partial \eta^2} \geq 0, \quad \text{for } \eta \geq 0$$

$$f(0, \gamma) = 0, \quad \frac{\partial f}{\partial \eta}(0) \leq k$$

In some ways this characterization of f runs counter to the intuition behind the Arrow-Pratt coefficient of risk aversion and some expected utility-derived theory of risk, which would predict f to be diminishing, i.e. $(\partial^2 f / \partial \eta^2) < 0$. However, the difference might be explained by other factors faced by firms and other actors in financial markets. Investment firms face some hard constraint on the total amount of risk that is deemed acceptable for them to carry at any given time. If one asset's inordinate riskiness begins to crowd out the firm's capacity to take on normal risk levels on other assets, this could be seen as imposing an additional cost and would explain the observed shape of f .

In our model, we can observe investors' rebalance frequency Φ for different critical risk thresholds η^* . Our goal is to relate the rebalancing frequency Φ to the proportional transactions cost k . We now define the link between the transactions cost and the risk threshold.

Consider the following function for f that displays the required characteristics.

$$(5) \quad f(\eta, \gamma) = \gamma\eta^2$$

From (4) and (5) we can write the threshold risk η^* as

$$(6) \quad \eta^* = \frac{k}{\gamma}$$

Thus, with this form for $f(\eta, \gamma)$, the effect of a change in η^* on rebalance frequency will be a scaling of the effect of k on the rebalancing frequency.

5. The Method

The B-S partial differential equation (PDE) can be solved in a number of ways. The numerical methods approach involves discretizing each element of the PDE into standard finite difference formulas. Once each element has been discretized into a difference formula, the formulas can be algebraically manipulated in order to solve for the entire solution space in a row-by-row, iterative fashion. For example, a discretization of the term V_S could be written

$$V_S \approx \frac{V(S + \Delta S, t) - V(S - \Delta S, t)}{2(\Delta S)}$$

and similarly for the other derivatives in the PDE, where ΔS is the size of the price-step. Then, by algebraic manipulation, the value of V at the next time step can be approximated using information found at nodes in the current timestep.

This method presents two challenges when used to solve the B-S PDE. First, unlike many equations solved using finite difference methods that run forward in time, we only know the value of the option at maturity, and thus must solve the equation backwards in time (a “terminal-value problem,” as opposed to the standard “initial value problem”). Second, the fact that one of the boundary conditions is itself a function of time poses challenges for the construction of the tri-diagonal multiplier matrix used to perform the iterative calculation. The general approach is called the Forward Difference method or Explicit Euler method.

An important aspect of a numerical method like the Euler is the possibility of instability. We are not concerned with the details of what causes the stability or instability in this method, and just note that the time steps must be sufficiently smaller than the space (underlying asset price) steps in order for the solution to be stable. There is an alternative method known as the Backward Difference or Implicit Euler, in which backward difference formulas are used for discretization and the tri-diagonal multiplier matrix is inverted prior to being multiplied by each row in the iterative solution process. While this method works in theory, a more complicated approach must be taken to deal with the two boundary conditions and involves the additional computation of inverting a matrix. The solution set uses as its terminal condition the call option payoff function

$$V(S, \tau = 0) = \max(S - E, 0)$$

where E is the option exercise price. This is the payoff function for a European call option¹. This type of option is very widely traded and requires simple computations when used in our analysis. Time is converted to run backwards using time to maturity τ . The boundary conditions (b.c.) on the B-S solution are

$$\begin{aligned} V(0, \tau) &= 0 \\ \lim_{S \rightarrow \infty} V(S, \tau) &= S - e^{-r\tau} \end{aligned}$$

For the numerical approximation we define S_{\max} as the largest price of the underlying asset, so that the b.c. become:

$$V(S, \tau) = \begin{cases} 0 & S = 0 \\ (S - Ee^{-r\tau}) & S = S_{\max} \end{cases}$$

Notice that the b.c. for $S = S_{\max}$ depends on time to maturity τ . The resulting numerical solution provides us with an approximation to the B-S option price V at each point in time τ and for each stock price S (given some volatility, riskless rate of return, and strike price)². This is significant in itself because it provides desired values for a wide range of circumstances in an efficient and complete manner. By contrast, accounting for such a complete range of circumstances using Monte Carlo methods takes a lot of time and repeated simulations, with no guarantee of success.

Additionally, the fact that we have an option price for each stock price and time (each node), allows us the possibility of traversing through a wide range of stock values in one time step, which is not possible in the binomial tree model. The most important result is that we can look at all possible paths including large stock price changes in small time steps. This is therefore an opportunity to make an original contribution to the understanding of the interaction between options trading and transactions costs.

6. The Model

Financial markets are opaque and exceedingly complex. It is not realistic to attempt to isolate the effect of a change in transactions costs in this setting. Our approach is to study a simplified simulation, akin to a controlled experiment.

We take from our numerical solution set a smaller matrix of option prices $V(S, \tau)$, on some interval of underlying asset price S and time to maturity τ , and then identify and map out every possible path through it. We then use this collection of paths to test the effect of changes in risk threshold η^* on the hedging behaviour of investors as they follow each path.

¹The B-S solutions for call and put options are symmetric. By call/put parity, an equivalent analysis of put options could be derived from our results.

²At this time, we are interested chiefly in the effect of changes in transactions costs on the frequency of options-related trading (such as delta hedging) and not specifically on its effect on the market price V of those options. Inclusion of transactions costs in the method of Leland or BV does indeed change V , but it has a negligible effect on the shape (ie. the partial derivatives) of the solution function $V(S, \tau)$ and it is these derivatives that influence the hedging behaviour of investors. As a result, we are satisfied for now to use a numerical solution for V that is built on the original Black-Scholes assumptions.

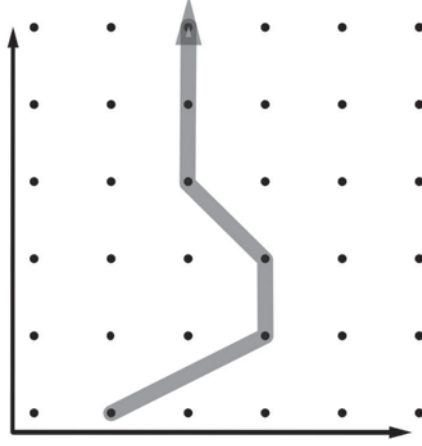


Figure 1: Example path through 6×6 matrix (with τ on the vertical axis and S on the horizontal axis)

The number of possible paths is a function of the dimensions (m, n) of the matrix we select for study. Specifically, the number of paths will be n^m , where m is the number of time-steps and n is the number of price-steps (steps in underlying asset price S).

At the beginning of each path, the investor sets up a portfolio with a long position in one call option, and $\Delta = V_S$ units of the underlying asset in a short position. At the next node in the path (S_i, τ_{m-1}) , the portfolio will be assessed for its hedging error and will be rebalanced if it satisfies (4). If not, the portfolio will maintain its previous hedging position into the next time-step, (S_i, τ_{m-2}) where all portfolios on all paths are again re-assessed. For the purpose of each assessment, hedging error ε will be defined as

$$(7) \quad \varepsilon = V_S(S_i, \tau_j) - \Delta$$

Where $V_S(S_i, \tau_j)$ is the value of V_S at the present node and Δ is the value of V_S at the past node where the portfolio hedging position was last rebalanced. Numerical approximations of the partial derivative V_S are obtained from the solution matrix using middle differences.

$$(8) \quad \hat{V}_S(S_i, \tau_j) = \frac{V_{i+1,j} - V_{i-1,j}}{2(\Delta S)}$$

Where $V_{i,j}$ is the option price, $\hat{V}_S(S_i, \tau_j)$ is the numerical estimate of its partial derivative with respect to S at node (S_i, τ_j) and ΔS is the size of the price-step.

The behaviour of the investor on each path (i.e., the number of portfolio rebalances) are observed for a range of values for the threshold η^* . Because of the exponential relationship between matrix size and number of paths, computing power is a limitation for our model. The largest matrix we can feasibly investigate is 6×6 . Although we have looked at a range of solutions, we present some quantitative analysis for one 6×6 matrix. In this case we used the following parameters in our B-S numerical solution

$$E = 50, \quad \sigma = 0.2, \quad r = 0.05, \quad \Delta S = 1, \quad \Delta \tau = 0.0001$$

Where E is the option strike price, σ is the standard deviation of the underlying asset price, r is the riskless interest rate, and ΔS and $\Delta \tau$ are the size of the price- and time-steps respectively. We take from this larger solution matrix the 6×6 square (such as the one in Fig.1) that surrounds the strike price at maturity and contains $6^6 = 46,656$ unique paths. Using the strategy outlined above, we can chart the effect of a change in risk threshold η^* on the total number of rebalances Φ that occur on all 46,656 paths during the time interval $\tau = (0, 6)$.

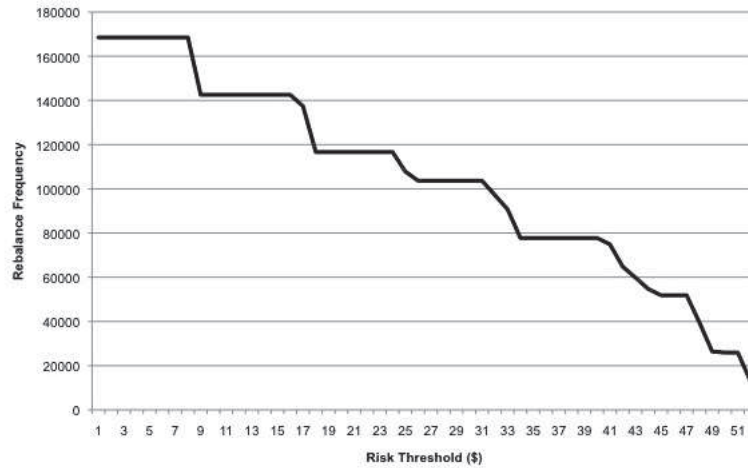


Figure 2: Risk threshold and rebalance frequency, for 6×6 matrix centred at strike price at maturity

Ideally, we would like to study larger matrices (again, computing power is the limiting factor). In the meantime we can move the small matrix further from maturity and/or further into or out of the money in order to gain more perspective on this relationship. The following figure illustrates a set of 5×5 matrices, each one a different distance from maturity (where darker paths are further from maturity). It is clear that the slope changes with time to maturity. This is because $\partial V_S / \partial \tau$ and $\partial V_S / \partial S$ are smaller when τ is larger.

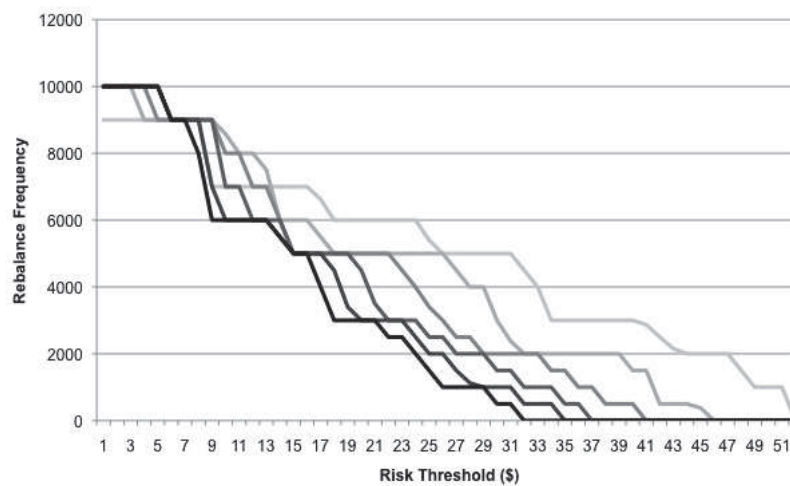


Figure 3: Risk threshold and rebalance frequency for a set of 5×5 matrices moving away from maturity

The fact that we are giving equal weight to every possible path through a matrix might suggest that we are giving a disproportionately high weight to a large number of unlikely price paths. Thus we examine

the characteristics of price variance that the set of observed price paths exhibit collectively. Specifically, we check whether our set of paths is consistent with an asset price standard deviation of $\sigma = 0.2$. We are able to measure volatility by counting the total price change along each path. It is difficult for us to extract from this information the price “variance” as it is strictly defined, but examining our proxy measure of volatility is still useful.

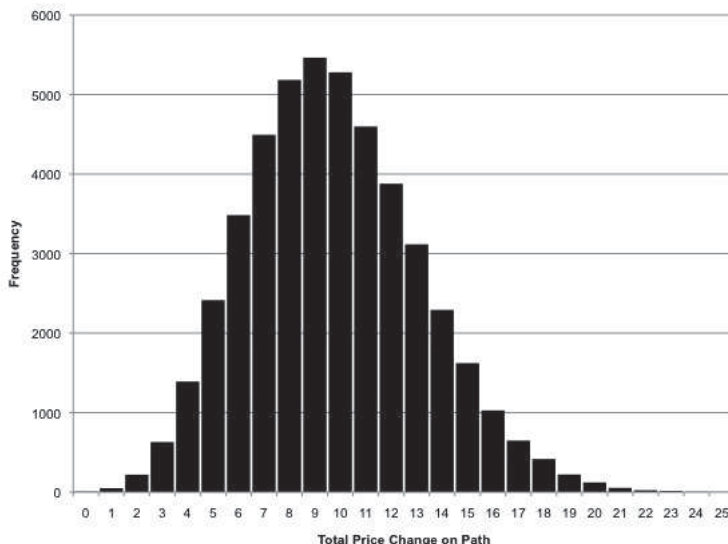


Figure 4: Price volatility frequency distribution

This distribution has the skewed shape expected given that the stochastic term in dS has a normal distribution, and so S has a log normal distribution. The most volatile path that can be taken through the matrix (there are two of them, mirror images of one another) with a total price change of 25, has an actual standard deviation of 2.7389. This is several orders of magnitude higher than the parameter value $\sigma = 0.2$, although the mean standard deviation of the whole set is lower than 2.7389³. This discrepancy could be rectified by using a matrix that is not square, but has more time-steps m than price-steps n . We could also use a smaller price-step ΔS or a larger value for the parameter σ .

7. Quantitative Estimation

We can use a simple OLS regression to fit a curve to the data. Although jagged, the relationship seems to be linear. We expect that using a larger matrix, with a sufficient number of steps in time and price, would yield a smooth relationship. We suggest the following specification⁴:

$$(9) \quad \widehat{\text{Frequency}}_i = \beta_0 + \beta_1 \text{Threshold}_i + \text{error}_i$$

where Frequency is the frequency with which portfolio hedging positions are rebalanced Φ , and Threshold is the risk threshold parameter η^* . We report the results in Table 1, with coefficient standard errors in parentheses.

³This is difficult to calculate precisely. Consider from fig.4 the mode volatility of 9. Paths with a wide range of standard deviations fall into this category, from a minimum of $\sigma \approx 0.95$ to a maximum of $\sigma \approx 2.35$. The mean value of σ for the entire set probably sits near the middle of this range.

⁴It is possible to achieve a marginally better fit by including a (negative) quadratic term in the model, but this disagrees with the slight convexity of some other results (as in Figure 3) and is not reported.

Estimate of (9)	
Obs.	52
β_0	180288.8 (2077.88)
β_1	-2852.15 (68.23)
$adj.R^2$	0.971

Table 1

The fit of this specification can be evaluated visually by comparing our prediction $\hat{\Phi}$ and the set of observations Φ_{obs} , as shown in Figure 5.

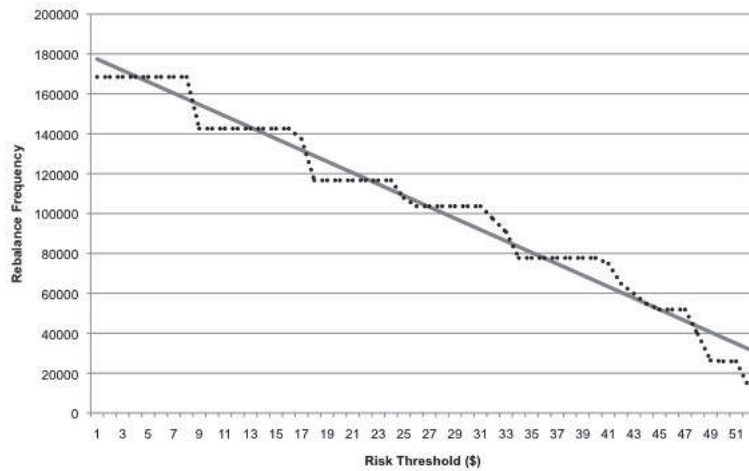


Figure 5: Fit of Model Estimation

Judging by the way it changes when the matrix is moved, the slope β_1 of this relationship depends on how close it is to the strike price and maturity. In other words, it seems to depend on the second partial derivatives $V_{S\tau}$ and V_{SS} , which is what we expect. The intercept β_0 , as a measure of the quantity of trading under the condition of no transactions costs, seems to be independent of these derivatives. Instead, it would depend on the number of agents in the market, and their technological capacity for high-frequency trading.

8. Conclusion and Further Research

Our model has allowed us to explore all the paths an option price can take within some bounds of asset price S and time to maturity τ (imposed by limitations on computing power⁵). From the 6×6 matrix we have been able to identify the nature of the effect of proportional transactions costs on the frequency of delta-hedge rebalancing. We are confident that what appears to be a jagged relationship would become smooth with sufficiently large number of time- and price-steps. Our quantitative analysis suggests the underlying relationship is linear, and can be written

⁵Recall that the number of paths through a matrix are a function m^n of its dimensions. As a consequence, the number of calculations required to process the analysis of these matrices grows exponentially with their size.

$$(10) \quad \Phi = \beta_0 - \beta_1 \left(\frac{k}{\gamma} \right)$$

where β_0 and β_1 are positive constants determined by market conditions, γ is a coefficient representing the risk aversion of investors and k is the proportional transactions cost. Observation of investor activity may give good estimates of these parameters. If these coefficients could be estimated then it would be possible to predict the effect of a change in transactions costs on the quantity of trading, at least in the part of the market that mimics the behaviour of the investors in our model⁶.

Besides investigating matrices that are larger than 6×6 , our method could be refined by including in the options pricing model the effect of changing transactions costs, possibly by the augmented variance method of Leland. We could also pursue the BV suggestion of endogenizing the trading frequency we have measured here into the option pricing model, through their version of Leland's augmented variance method. Currently the option price solution set used in our model is constructed using the original B-S PDE, and the transactions costs are not reflected in those prices. Finally, we could adjust the parameters of our analysis to be consistent with observed asset price volatility.

⁶It would also be necessary to account for any features of the market such as fixed transactions costs or limits on the divisibility of assets, that might otherwise affect an investor's band of inaction and how it interacts with the proportional transactions cost k .

References

Black, F. and Scholes, M. *The pricing of options and corporate liabilities*, Journal of Political Economy, 3 1973 pp 637-654, The University of Chicago Press.

Boyle, P. and Vorst, T. *Option Replication in Discrete Time with Transactions Costs*, The Journal of Finance, Mar. 1992, pp 271-293, American Finance Association.

Cox et al. *Option Pricing: A simplified approach*, Journal of Financial Economics, Sep. 1979, pp 229-263, Elsevier.

Leland, Hayne E. *Option Pricing and Replication with transactions Costs*, The Journal of Finance, Dec. 1985, pp 1283-1301, American Finance Association.

Pratt, John W. *Risk Aversion in the Small and in the Large*, Econometrica, Jan.-Apr. 1964, pp 122-136, The Econometric Society.

ASB 2013
Antigonish, Nova Scotia

George Ye
Christine Panasian
Sobey School of Business
Saint Mary's University

HOW MANY DAYS EQUAL A YEAR? A NOTE ON THE MEAN-VARIANCE MODEL

In portfolio management, the means and variances of stock returns are usually estimated on a daily basis and then converted to longer periods of time. This paper examines the issue of how to convert 1-day means for longer periods and investigates the impacts of this conversion on capital allocation decisions and portfolio performance evaluations.

1. Introduction

The mean-variance model and the related efficient frontier provided the basis of the development of modern portfolio theory. It offered the first systematic analysis of the risk-return trade-off faced by each investor and it still constitutes a fundamental theory for portfolio formation. Markowitz's ground-breaking theoretical model and associated numerical solutions of portfolio selection under uncertainty supported the foundation for many important developments in financial economics like the Capital Asset Pricing Model (CAPM), derived by Sharpe (1964), Lintner (1965), and Mossin (1966)) and the now classical distinction between systematic and idiosyncratic risks. Although the importance of his work remains undeniable, mean-variance portfolio analysis generated more questions than answers and, consequently, caused a large body of research.

In this paper, we address the importance of an issue which is very practical, but often ignored, on estimating the means and variances of stock returns. In practice, mean and variance on a stock return is often estimated on a 1-day horizon, and then they are converted to longer horizons as needed, by scaling. However, this conversion is not always trivial, because the time scale used for this conversion may not be the same as the calendar days and depends on the stock price movement.

First, in the literature, Fama (1965), French (1980), Roll (1984), French and Roll (1986) and many other researchers find that volatility is caused by trading activities. This means that only trading days should be counted to determine the time scale. For example, the 1-year variance of return on a stock traded on the New York Stock Exchange (NYSE) should be

equal to 1-day variance multiplied by 252, the number of trading days in a year on NYSE, rather than by 365, as shown in Hull (2012). Although some researchers, including Smithsonite and Minton (1996 a,b), J.P. Morgan (1996), Diebold et al (1997), among others, show different perspectives, this approach has been widely accepted by industry and regulators such as the Basel committee.

However, little attention has been paid to the issue of converting a 1-day mean to a longer time horizon. While the volatility is caused by trading activities, the mean of a series of returns depends on a variety of factors among which time and the time value of money consideration. This implies that non-trading days should also be counted toward calculating the time scale. At the same time, since only trading days contribute to volatility, or market price risk, the expected returns on no trading days cannot be same as those on trading days. Thus, for instance, the simple rule of multiplying by 252 may not be correct to convert a 1-day mean to 1-year mean.

In this paper, we present an approach to convert a 1-day mean to a longer horizon, based on the Capital Asset Pricing Model (CAPM), and show how this conversion may affect the decisions of capital allocation and on the portfolio performance evaluation.

2. Mathematical Framework

Let $E(r_x)$ and σ_x be the mean and standard deviation of the daily return on stock X, i.e., 1-day mean and 1-day standard deviation, respectively. If the time horizon of our investment is a period of N calendar days, in order to use the mean-variance model to make optimal investment decisions, we need to adjust these parameters to fit our time horizon.

We denote $E^*(r_x)$ and σ_x^* to be the mean and standard deviation of the return on stock X for the time period of N calendar days, and n the number of trading days in the period. Then, since only trading days contribute to volatility, we have:

$$\sigma_x^* = \sigma_x \sqrt{n} \quad (1)$$

However, not only the number of trading days contributes to the determination of the mean, but also that of the non-trading days. Thus, the mean for an N-day period is consisting of two components: the total mean of n trading days and the total mean of (N-n) non-trading days. The first component is the 1-day mean multiplied by the number of trading days, i.e., $nE(r_x)$. Regarding the second component, in the light of the CAPM, the expected return consists of the risk free rate of return and a risk premium, representing compensation for investment time and for the risk assumed, respectively. During a non-trading day, an investment will require compensation for time value of money only. However, since risk or volatility is mainly caused by trading activities, the risk on a non-trading day can be ignored, and the risk premium required should be close to zero. Thus, the second component is equal to $(N-n)R_F$, where R_F is the daily risk free rate of interest. Therefore, in summary, we can calculate the mean for an N-day period as follows:

$$E^*(r_x) = nE(r_x) + (N-n)R_F \quad (2)$$

In some financial markets, such as foreign currency exchange market, trading is carried out seven days a week, and therefore N , the number of calendar days, is the same as n , the number of trading days. Since both the mean and standard deviation are adjusted by the same time scale, the investment horizon should have no impact on the portfolio decision.

However, in other financial markets, such as stock markets around the globe, trading is interrupted during weekends and stock market holidays, and therefore, number of trading days is less than the number of calendar days. For instance, there are 252 trading days on the New York Stock Exchange in one year, that is $n=252$ while there are $N=365$ calendar days. Thus, formula to convert 1-day mean and standard deviation to 1-year ones are:

$$E^*(r_x) = 252E(r_x) + 113R_F \quad (3)$$

$$\sigma_x^* = \sigma_x \sqrt{252} \quad (4)$$

Since the mean and standard deviation are adjusted by different time scales, the optimal portfolio selection may differ with different investment horizon. In next section, we will show some examples on this effect.

3. Effects of Time Horizon Changes

3.1. Effects of on Capital Allocation

For the purpose of illustration, we present a numerical example as follows: From the historical daily data from December 1, 2009, to December 1, 2012, we obtained the estimates of the mean and standard deviation of the daily return on the S&P 500 as below:

$$E(r_{S\&P500}) = 0.000407$$

$$\sigma_{S\&P500} = 0.011648$$

For simplicity assume the risk free rate is $R_f^* = 3\%$ per annum, or $R_f = 0.000082$ per day.¹

Following Bodie et al (2012), we assume an investor has a utility function as follows:

$$U = E(r) - (A/2) \sigma^2$$

with a degree of risk aversion, A , equal to 4,

First, assuming that the investment horizon is one day, following Bodie et al. (2012),

¹ $0.03/365 = 0.000082$

the optimal capital allocation between the risk free asset and the equity for the investor is:

$$y_d = (E(r_{S\&P500}) - R_F) / (A \sigma_{S\&P500}^2) = (0.000407 - 0.000082) / (4 \times 0.011648^2) = 60\%$$

This suggests investing 60% in the equity portfolio and the remaining 40% in T-Bills.

Now, assuming that the investment horizon is one year and using Eq. (3) and (4), the annual mean and standard deviations are calculated as follows:

$$E^*(r_{S\&P500}) = 252 \times 0.000407 + 113 \times 0.000018 = 0.1046$$

$$\sigma^*_{S\&P500} = 0.011648 \times 252^{1/2} = 0.1849$$

Similarly, by following Bodie et al. (2012) the optimal allocation between the risk free asset and the equity is:

$$\begin{aligned} y_a &= (E^*(r_{S\&P500}) - R^*_f) / (A \sigma^*_{S\&P500})^2 \\ &= (0.1046 - 0.03) / (4 \times 0.1849^2) \\ &= 55\% \end{aligned}$$

This suggests investing 55% in the equity portfolio and the remaining 45% in T-Bills.

The above simple example shows that the investment horizon could affect capital allocation decision considerably.

3.2. Effect on Performance Measurement

On a daily basis, the Sharpe ratio of an asset is given by:

$$S_x = (E(r_x) - R_F) / \sigma_x$$

and on an annual basis, the Sharpe ratio is

$$S^*_x = (E^*(r_x) - R^*_F) / \sigma^*_x$$

Since

$$E^*(r_x) - R^*_F = nE(r_x) + (N-n)R_F - NR_F = n(E(r_x) - R_F)$$

the annual Sharpe ratio becomes:

$$\begin{aligned} S^*_x &= n(E(r_x) - R_F) / \sigma_x n^{1/2} \\ &= n^{1/2} S_x \end{aligned}$$

The above equation indicates that, though the value of the Sharpe ratio is dependent on the

investment horizon, the ranking among securities traded in the same market will remain the same. For example, for two portfolios of NYSE traded stocks, if one is better than the other based on the daily Sharpe ratio, the same ranking will be maintained if the annual Sharpe ratio is used. However, when we compare the performances of securities traded in different markets with different number of trading days, the investment horizon may affect their ranking.

For the purpose of illustration, we can look at the following numerical example:

Consider a stock X traded on NYSE and a foreign currency Y, traded on the FX market. Assume that the means and standard deviations of the annual rate of returns on X and Y are given as follows:

$$\text{Stock X: } E^*(r_X) = 20\%, \sigma^*_{X} = 30\%$$

$$\text{Currency Y: } E^*(r_Y) = 12\%, \sigma^*_{Y} = 8\%$$

The risk free interest rate is $R^*_F = 4\%$ per annum.

The number of calendar days is $N=365$, while the number of trading days for X is $n_x=252$ and that for Y is $n_y=365$. Then on an annual basis, the Sharpe ratios of X and Y are as follows:

$$S^*_X = (E^*(r_X) - R^*_F) / \sigma^*_{X} = (0.2 - 0.04) / 0.3 = 0.53$$

$$S^*_Y = (E^*(r_Y) - R^*_F) / \sigma^*_{Y} = (0.12 - 0.04) / 0.14 = 0.57$$

Since $S^*_X < S^*_Y$, we conclude that currency Y has better performance than stock X.

If we perform the same analysis on a daily basis, using equations (2) and (3), the means and standard deviations of the daily rate of returns on X and Y are calculated as follows:

$$\text{Stock X: } E(r_X) = 0.000745, \sigma_X = 0.018898$$

$$\text{Currency Y: } E(r_Y) = 0.000329, \sigma_Y = 0.007328$$

The risk free interest rate is $R_F = 0.000110$ per day.

Then the Sharpe ratios of X and Y are as follows:

$$S_X = (0.000745 - 0.000110) / 0.018898 = 0.0336$$

$$S_Y = (0.000329 - 0.000110) / 0.007328 = 0.0299$$

Since $S_X > S_Y$, we conclude that stock X has better performance than currency Y.

The above example indicates that the difference in investment horizon may affect the performance ranking of the assets.

One note worth mentioning in this context is that since beta is independent of time

horizon, investment horizon should not affect the performance ranking given by the beta-based performance measures, such as Treynor ratio.

4. Conclusions and Suggestions

When using the mean-variance framework in portfolio management, it is common to estimate the 1-day means and variances of stocks and other assets, and then convert them to ones with longer horizons, such as annual means and variance. While the conversion of variances has been widely addressed and there is an approach popularly accepted, the conversion of means has been rarely addressed.

In this paper, we present a formula to convert 1-day means to 1-year means or means with other horizons, based on the CAPM model. Then, with this formula, we investigate the impact of time horizon changes to the optimal capital allocation decision and to the performance evaluation, with numerical examples.

We show that the optimal capital allocation decision can be different for a particular investor if he or she would change his or her investment horizon. In addition, when the Sharpe ratio is used to evaluate the performance of portfolios, the choice of the time horizon may significantly affect the evaluation.

While we present a simple approach for the conversion of means with different time horizons, the issue is much more complex. As we are working in the mean-variance framework, the expected return consists of 2 distinguishable components: the risk-free rate and market risk premium. Thus, since market risk is mainly caused by trading activities, risk premium on non-trading days can be safely ignored. Consequently, the mean on a non-trading day should include only the risk-free interest rate. However, many recent studies have shown that there are some other factors which also contribute to mean levels. One of the most significant developments in this direction of researches is the Liquidity-based Asset Pricing Model (LAPM) developed by Amihud and Mendelson (1986), Acharya and Pedersen (2005), and Liu (2006), among others, which indicates that illiquidity is priced as a security characteristic and/or as a risk factor and such, the illiquidity premium should be added to the expected returns. In the light of the LAPM model, non-trading days should have higher illiquidity premium than trading days, although they have lower market risk premiums. How to incorporate these issues into the conversion formula for means will be a direction of our future research.

References

Acharya, V. and Pedersen, L., "Asset Pricing with Liquidity Risk," *Journal of Financial Economics*, 77, (2005), 375–410.

- Amihud, Y. and Mendelson, H., “Asset Pricing and the Bid–Ask Spread,” *Journal of Financial Economics*, 17, (1986), 223–249.
- Bodie, Z., Kane, A. and Markus, A., *Investments*, 9th ed, McGraw-Hill Ryerson, 2011.
- Diebold, F. X., Hickman, A., Inoue, A., & Schuermann, T., “Converting 1-day Volatility to h-day Volatility: Scaling by h Is Worse Than You Think,” *University of Pennsylvania working paper*, (1997).
- Fama, E., “The Behavior of Stock-Market Prices,” *Journal of Business*, 38(1), (1965), 34-105.
- French, Kenneth R, “Stock Returns and the Weekend Effect,” *Journal of Financial Economics*, 8, (1980), 55-69.
- French, K. and Roll, R., “Stock Return Variances: The Arrival of Information and the Reaction of Traders,” *Journal of Financial Economics*, 17, (1986), 5-26.
- Hull, J., *Options, Futures, and Other Derivatives*, 8th ed., *Prentice Hall*: (2012).
- J.P. Morgan, *RiskMetrics—Technical Document*, 4th ed., New York, (1996).
- Lintner, John, “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets,” *Review of Economics and Statistics*, 47, (1965), 13-37.
- Liu, W., “A Liquidity-Augmented Capital Asset Pricing Model,” *Journal of Financial Economics*, 82, (2006), 631–671.
- Mossin, Jan, “Equilibrium in a Capital Asset Market,” *Econometrica*, 35, (1966), 768-783.
- Roll, R., “Orange Juice and Weather,” *American Economic Review*, 74 (5), (1984), 861-880.
- Sharpe, William F., “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” *Journal of Finance*, 19, (1964), 425-442.
- Smithson, C. and Minton, L., “Value at Risk” , *Risk*, 9, (January, 1996a).
- Smithson, C. and Minton, L., “Value at Risk (2)” , *Risk*, 9, (February, 1996b).

ASB 2013
Antigonish, Nova Scotia

Imran A. Tasadduq
Department of Computer Engineering
Umm Al-Qura University, Saudi Arabia

M.H. Imam
Department of Civil Engineering
Umm Al-Qura University, Saudi Arabia

Abdul-Rahim Ahmad
Faculty of Management
Rowe School of Business
Dalhousie University, Canada

A NOVEL ADAPTIVE BOUNDARY SEARCH ALGORITHM FOR SOLVING FACILITY LAYOUT PROBLEMS¹

An highly effective algorithm is presented to solve open space facilities layout planning problems involving modules with constant aspect ratios. This algorithm is a novel combination of steepest descent and corner search. Layout solutions produced by the proposed algorithm for well-known test problems were found superior to the published layouts as well as the best layouts produced by commercial layout design software VIP-PLANOPT.

Introduction

Facility layout planning problem is a well-studied combinatorial optimization problem arising in a variety of applications (Drira, 2007; Taghavi, 2011; Anjos, 2010; Yang, 2011; Saif, 2004). Most commonly, the goal is to produce high-quality optimal layouts for problems involving rectangular blocks or modules representing functional units like department, machine, room, cell or empty space. In facilities layout planning problems, locations of a given number of such modules are determined to minimize a specific cost function such that no overlapping exists between the modules. This problem is a known NP-complete problem (Sahni, 1976) and therefore, in almost all cases, a verifiably optimal solution is not known. Hence, all published work in this area has presented the performance of new algorithms comparing them with the previously published algorithms. A review of different approaches reported in the literature to solve facility layout planning problems is given in (Singh, 2006).

¹ The authors would like to acknowledge the generous support and resources extended by Umm Al-Qura University, University of Waterloo, and Dalhousie University throughout the course of this research. This research has been partially supported by a research grant from the Faculty of Management, Dalhousie University, Canada.

There are different types of facility layout planning problems. Facility layout planning involving modules of equal area is referred to as cell assignment problem. In such problems, a certain area is divided into identical cells and each facility is assigned to one of the cells. Several approaches have been used. Recently, (Ramkumar, 2009) have solved the problem by treating it as Quadratic Assignment Problem and (James, 2009) present an example of solving such problems using Tabu search. Other methods for solving these problems may be found in (Raoot, 1994; Ahuja, 2000; Misevicius, 2003; Talbi, 2001).

In case of facility layout planning involving modules of non-identical areas, the problem, the methods developed for solving equal area facilities layout problem cannot be used (Kado, 1995). These problems involving unequal area modules can be classified into two sub-groups considering the modules as “Soft” or “Hard”. A soft module has pre-specified area of variable aspect ratio (defined as the ratio of the vertical length of the facility to its horizontal length) with given upper and lower bounds on them. There are several published articles on such problems that involve only soft modules. Some of the most recent articles in this area include (Komarudin, 2010; McKendall, 2010; Scholz, 2010; Scholz, 2009).

In contrast with soft module unequal area facility layout planning, there are problems that involve only “Hard” modules. A hard module has fixed dimensions with a constant aspect ratio (VIP-Planopt, 2011). In many applications, the solution to such problems is required because the modules have pre-specified dimensions to suit the particular applications. Also, the flexibility of varying the aspect ratios does not exist. Such problems are harder to solve and therefore relatively fewer publications report techniques for solving such problems. Nevertheless, this class of problem is of particular interest due to its value to both researchers and practitioners in facility planning domain.

A facility layout may be static or dynamic. Static facility layout problem is the one where a given cost function is optimized and material flows between modules are assumed to be constant over the planning horizon. When material flows between departments change during the planning horizon, the problem becomes a dynamic facility layout planning problem. Several approaches have been used to solve these problems, for example, (McKendall, 2010; Sahin, 2010; Dong, 2009; Ulutas, 2009). However, for both static and dynamic facility layout planning, the basic optimization remains the same and an efficient technique developed for static layout planning may be incorporated in a dynamic layout planning, as done by (McKendall, 2010) who used the technique developed in (Imam, 1998) to solve the dynamic facility layout planning problem.

Facility layout planning problems may also be classified based on existence or absence of the enclosing area constraint. The enclosing area constraint may be imposed to obtain the optimal layouts with a given rectangular, a single row (Samarghandi, 2010; Anjos, 2009), or multiple rows (Ficko, 2010). On the other hand, facility layout problems with open space (Mir, 2001) do not impose any constraint on the enclosing area.

The objective of this paper is to present a novel algorithm for solving the hard module unequal area facility layout planning problem in an open space. Over the past decade, little research has been done to solve this class of problems.

Hard module unequal area facility layout planning problems have been solved using analytical methods and other techniques including evolutionary algorithms. Among analytical methods, seven algorithms have been reported in (Imam, 1989; Mir, 1990; Mir, 1992; Imam, 1993; Mir, 1996; Imam, 1998; Mir, 2001). In one of these publications, a hybrid optimization

approach is used that combines Simulated Annealing and Analytical Method (Mir, 2001). In another publication, the idea of Cluster Boundary Tracking is presented with remarkable results (Imam, 1998). The same authors have reported a number of other algorithms, such as, analytic annealing (Mir, 1996), optimization through controlled convergence (Imam, 1993), and nonlinear programming (Imam, 1989). Recently, (McKendall, 2010) have used the analytical method of Cluster Boundary Tracking method published by (Imam, 1998) the most recent paper that has dealt with these types of problems.

It is observed that since 2001, analytical methods have not been published for solving these types of problems and the thrust has been on evolutionary algorithms. However, examples of heuristics and evolutionary algorithms are very few. Simulated Annealing has been successfully applied to by (Tam, 1992), evolutionary algorithms used for solving these problems mainly cover Genetic Algorithms (GA). A good survey of some of the earlier work that addresses hard module unequal area facility layout planning using GAs has been reported in (Kado, 1995). Although there are several other papers that have used GAs to solve facility layout planning problems, such research focused on other types of facility layout planning problems such as those involving equal-area modules, soft modules, etc.

This paper presents an effective algorithm for solving the special class of open-space facility layout planning involving hard modules of unequal dimensions. The proposed algorithm is a novel combination of heuristic cluster boundary search construction phase and analytic steepest descent improvement phase of optimization. It is to be noted that the analytic steepest descent for layout improvement has been employed in past (Imam, 1989; Mir, 1990). In addition, the heuristic cluster boundary search using a corner search has also been reported (Ahmad, 2006; Ahmad, 2005; Ahmad, 2005; Welgama, 1993). However, little research can be found on synergistically employing the two techniques in one algorithm, a venue that offers significant promise. Simulation results presented in this paper demonstrate that the proposed technique is more effective than all previously published algorithms for open-space facility layout planning involving hard modules of unequal dimensions

The rest of the paper is organized as follows. Section II provides the problem statement. Section III describes benchmark problems used to test the proposed algorithm. Section IV provides detailed description of the proposed algorithm. Section V provides results and discussions. Section VI concludes with some interesting research directions.

Problem Statement

Let there be N modules of fixed dimensions to be placed at optimal positions in the Euclidean plane without any overlaps such that a given cost function C is minimized. The position of a module i is denoted by the coordinates of its centroid (x_i, y_i) . Let (L_i, W_i) denote the length and width of module i along the X- and Y-axes, respectively. If f_{ij} represents the cost of flow per unit distance between modules i and j and d_{ij} represents the inter-module distance measured between the centroid of modules i and j then the cost function C is defined as follows:

$$\begin{aligned} & \text{Minimize } C(x_1, y_1, x_2, y_2, \dots, x_N, y_N) \\ & = \sum_{i=1}^{N-1} \sum_{j=i+1}^N f_{ij} d_{ij} \end{aligned} \quad (1)$$

The overlap area A_{ij} between two modules i and j is defined in the following using (Mir, 2001):

$$A_{ij} = \lambda_{ij} (\Delta X_{ij}) (\Delta Y_{ij}) \quad (2)$$

where

$$\Delta X_{ij} = \left(\frac{L_i + L_j}{2} \right) - |x_i - x_j| \quad (3)$$

$$\Delta Y_{ij} = \left(\frac{W_i + W_j}{2} \right) - |y_i - y_j| \quad (4)$$

$$\lambda_{ij} = \begin{cases} -1 & \text{for } \Delta X_{ij} \leq 0 \text{ and } \Delta Y_{ij} \leq 0 \\ +1 & \text{otherwise} \end{cases} \quad (5)$$

A positive value of A_{ij} indicates there is an overlap between modules i and j . To avoid overlapping of the modules, the A_{ij} is restricted to zero. In the light of this discussion and Equations (1) to (4), the open-space, unequal area and hard module layout planning problem can now be formulated as:

$$\begin{aligned} & \text{Minimize } C(x_1, y_1, x_2, y_2, \dots, x_N, y_N) \\ & \text{subject to} \quad A_{ij} \leq 0; \quad i = 1, 2, \dots, N-1; j = i+1 \text{ to } N \end{aligned} \quad (6)$$

Benchmark Problems

In Table I, we provide a tabular compilation of three relatively large-scale published benchmark problems that we used to test our algorithm. In this table, each row contains a separate problem with columns giving the serial number, the problem ID along with the first known citation of the problem, the problem size in terms of the number of modules, the distance norm used and the cost function used.

Proposed Algorithm

We propose a hybrid algorithm involving dual hybridizations: construction-cum-improvement hybrid and analytic-cum-heuristic hybrid. We call the proposed algorithm as Adaptive Boundary Search (ABS) algorithm. The algorithm employs a Near-Optimality Hypothesis and a definition of Local Optimum Layout for reducing the otherwise infinite search space. Both of these concepts are described in the following.

Table I: Benchmark Problems Used to Test the Algorithm

	Problem ID [Original Source]	Size (Modules)	Distance Norm	Cost Function
1	OUH-020-IMx (Imam, 1993)	20	Rectilinear	$\sum_{i=1}^{N-1} \sum_{j=i+1}^N f_{ij} d_{ij}$
2	OUH-028-MIx (Mir, 1992)	28	Squared Euclidean	$\sum_{i=1}^{N-1} \sum_{j=i+1}^N f_{ij} d_{ij}$
3	OUH-050-EOS (VIP-Planopt, 2012)	50	Euclidean	$\sum_{i=1}^{N-1} \sum_{j=i+1}^N f_{ij} d_{ij}$

The traditional definition of local optima of satisfying the stationary conditions cannot be used in facility layout planning because of the discontinuities in the cost function occurring due to the overlap constraints (Imam, 1989; Mir, 1989). We define Local Optimum Layout as one in which cost cannot be improved by slightly perturbing the position of any module taken one at a time.

We state our Near-Optimality Hypothesis as follows. In a constructive-cum-improvement algorithm, modules are added to the layout one by one followed by an improvement cycle after every addition of module. If a Local Optimal Layout is achieved at each stage of optimization then the final layout will be a near optimal layout.

Here we provide a detailed description of the main components of the proposed algorithm. We also provide the pseudo-code of our algorithm in Appendix A.

The algorithm starts with an initial layout of two modules by randomly selecting a module as the first Incoming Module (IM1) and placing it at an arbitrary position. Once placed, this module will be called Placed Module (PM). This is followed by randomly selecting another module as the second Incoming Module (IM2) and placing it optimally on the boundary of the PM resulting in the Initial Boundary Placement (IBP), inspired by the idea described in (Imam, 1998). At this stage, the partial layout is also the global optimum layout and there is no need to perform an improvement cycle until there are more than two PMs.

The main loop in the ABS algorithm loop starts with the selection of the third module. In each iteration of this loop, an incoming module IM is randomly selected from the list of candidate modules followed by IBP to determine the best possible position on the boundary of the partial layout of PMs. The algorithm merges any mini-clusters formed during placements. After the placement of the IM through IBP, the improvement cycle starts. Each module of the cluster of PMs is moved one by one to obtain further gains in the cost through steepest descent. We refer to this procedure as Analytic Improvement, which is repeated until no gain is possible by moving any of the PMs. At this stage, the layout is a Local Optimal Layout. Subsequently, another module is randomly selected as an IM from the list of unplaced modules. The IBP is performed and the process to achieve a Local Optimal Layout through analytic improvement is repeated. The loop continues until all modules have been placed.

The Heuristic Boundary Search we have used explores positioning of an incoming module only at vacant corners of the placed modules. This heuristic search reduces the otherwise infinite solution space to tractable limits. Since our placement decisions are restricted only to the corners of placed modules, for a given module, there are only $O(n)$ possible locations supporting the use of some fast pseudo-exhaustive search in a hierarchical manner (Ahmad, 2006; Ahmad, 2005; Welgama, 1993; Imam, 1998).

As shown in the pseudo-code given in Appendix A, the Heuristic Boundary Search proceeds by investigating the placement opportunities for each of the four corners of an IM (or PM_j , as the case may be) at each of the four corners of the placed module PM_k . Given i modules in the cluster, there are a maximum of $16i$ possible locations to explore, a very loose upper bound. In the proposed algorithm, the Heuristic Boundary Search is called twice, first to place every incoming module (IM) and then to find a better location for a placed module (PM_j) after the application of the steepest descent improvement technique.

Once the best position for an incoming module has been found on the cluster of placed modules using Heuristic Boundary Search, the analytic improvement through steepest descent is carried out on all placed modules. For this purpose, the slopes of steepest descent for all placed modules are determined.

Following the decreasing order of slopes, each PM is moved in the direction of the steepest descent. The slope of steepest descent S_p for the p th module from the list of PMs is given by the following equation:

$$S_p = \sqrt{\left(\frac{\partial C}{\partial x}\right)^2 + \left(\frac{\partial C}{\partial y}\right)^2}$$

Whereas, the direction of steepest descent is given by the vector $\left\{\left(\frac{\partial C}{\partial x}\right) \mathbf{i}, \left(\frac{\partial C}{\partial y}\right) \mathbf{j}\right\}$

This analytic improvement procedure is explained through the example given in Figure 1, showing a cluster of eight PMs formed after applying Heuristic Boundary Search on IM8. Since the best position of IM8 on the boundary of the cluster has been obtained, analytic improvement through steepest descent procedure will be applied to each of the PMs.

In the example shown in Figure 1, module 6 has the highest value of SP . This module is therefore moved first in its direction of steepest descent, which is vertically upwards. The optimal position is shown in Figure 1(b), while ignoring the overlap with Module 3. To remove any overlap, heuristic boundary search is applied to Module 3 for determining its best position on the boundary of the cluster, resulting in a layout as shown in Figure 1(c). It is to be noted that the analytic improvement may result in an overlap of multiple modules. In such a case, all overlapping modules, except the module moved by the analytic improvement procedure, are placed at superior positions using heuristic boundary search.

If the new layout does not result in an improved cost, the improvement procedure is backtracked and the module with the next highest slope is moved and the procedure is repeated. If the move is successful, the slopes are calculated again and the analytic improvement procedure is repeated. Once all modules have been considered, or no further improvement is possible, the heuristic boundary search is applied to all modules. This cycle of analytic improvement

through steepest descent and heuristic boundary search is repeated until no gains are possible and a Local Optimal Layout is achieved.

Our heuristic improvement is inspired by ideas explored in (Imam, 2010). It involves minor horizontal and vertical perturbations in the position of each module present in the cluster, by considering one module at a time. These perturbations are retained only when these result in cost improvement. Furthermore, to reduce computational complexity we perform these heuristic improvements in a slightly greedy manner. For instance, a useful rightward perturbation would mean ignoring any leftward perturbation or, vice versa, but vertical perturbations are still explored. This final tuning was performed only once at the end of all iterations of analytic and heuristic

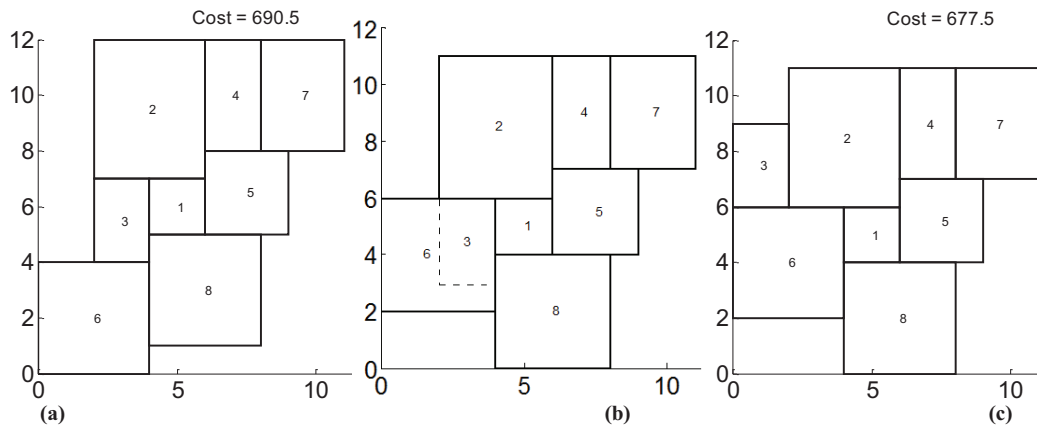


Figure 1: Example of application of the analytic improvement procedure
(a) Before analytic improvement (b) Steepest descent move of Module 6 (c) Module 3 moved to optimal boundary location

improvements. It involves detecting any super modules, where adjacent modules perfectly align along one side either vertically or horizontally.

Results and Discussion

The proposed algorithm was implemented using MATLAB and was used to solve the benchmark problems listed in Table I using a 2.0 GHz quad-core computer running on Windows 7 with a 8 GB main memory. The data for the benchmark problems can be found in (Tasadduq, 2012). Comparison of layouts obtained from the proposed ABS algorithm with the verifiably best available published results and those obtained from VIP-PLANOPT (VIP-Planopt, 2012) – an extremely effective commercially available layout optimization software – are presented here (Heragu, 2006; Tompkins, 2002). Table II provides a summary of these comparative analyses, clearly indicating the superiority of the proposed algorithm. Detailed discussions on these comparative analyses are provided here.

Problem OUH-020-IMx

This twenty-module problem was first introduced in (Imam, 1993) and a solution with a cost of 1,264.94. Subsequently, the problem has been solved by (McKendall, 2010) reporting a solution with a cost of 1151.40. We were able to obtain and verify the layout data from the authors through personal communications. Recently, this problem has been solved using a Memetic Algorithm resulting in a superior solution with a cost of 1,199.50 (Tasadduq, 2011). The best solution produced by VIP-PLANOPT has a cost of 1,157.00 (VIP-Planopt, 2012). In comparison, the proposed ABS algorithm yields the best layout with a cost of 1,140.00, shown in Figure 2, which is the best solution reported for any algorithm thus far.

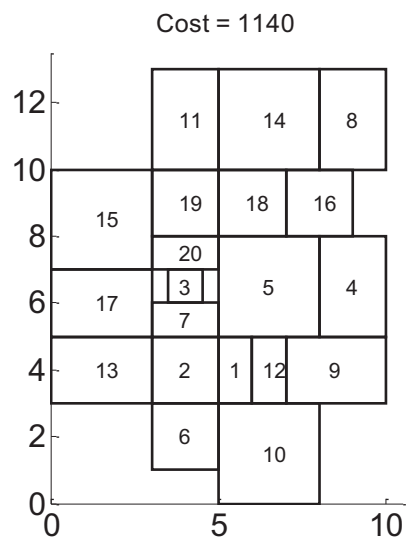


Figure 2: Optimal layout for problem OUH-020-IMx

Problem OUH-028-MIx

This randomly generated 28-module problem was presented as a test problem in (Mir, 1992). The paper reports a layout with a cost of 9,626.00 and provides all the data for verification. Later, it was used as a test problem for a new technique called Analytical Annealing, reporting a cost of 8,640.00 for the best layout obtained (Mir, 1996). Complete data were provided to verify the cost of the layout. The best layout produced by VIP-PLANOPT for this problem has a cost of 6,447.25 (VIP-Planopt, 2012). Our proposed ABS technique yields a cost of 6,289.64 for the best layout shown in Figure 3, which is significantly better than the previous best solutions as well as the one obtained by VIP-PLANOPT.

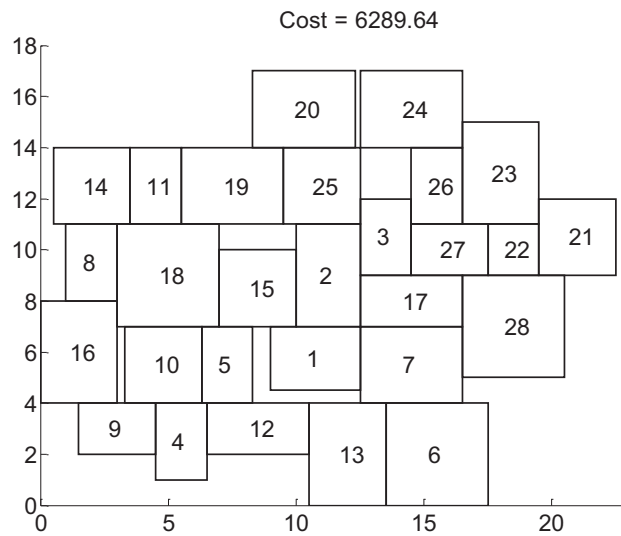


Figure 3: Optimal layout for problem OUH-028-Mix

Problem OUH-050-EOS

This 50-module problem was introduced by VIP-PLANOPT as a randomly generated benchmark problem in 1996 (VIP-Planopt, 2012). Previously, no test or benchmark problem of 50 or more modules was presented in the literature on facility layout planning. The best layout obtained by VIP-PLANOPT has a cost of 78,224.70 for this problem (VIP-Planopt, 2012). Subsequently, it was used as test problem in (Mir, 2001) for their hybrid search algorithm. They obtained a layout with a cost of 80,794.24. Recently, it has been used as a test problem in (McKendall, 2010) and the reported cost of the best solution as 71,291.40. Although no layout has been shown in the paper and no data are given to verify the given cost, we were able to obtain and verify the layout data from the authors through personal communication. This 50-module problem was also used in (Tasadduq, 2011) as a test problem and the reported layout cost was 77,504.00. The proposed ABS technique is able to obtain a layout with a cost of 71,151.41, which is lower than the reported cost by any published source or commercial software. The layout is shown in Figure 4.

Summary of Results

A comparison of the obtained results is summarized in Table II. The results clearly demonstrate that proposed technique is more effective than all previously published algorithms as well as popular commercially available layout optimization software. Data relevant to the layouts optimized using the proposed algorithm, such as coordinates of the lower left corners and problem dimensions, etc. are available from (Tasadduq, 2012).

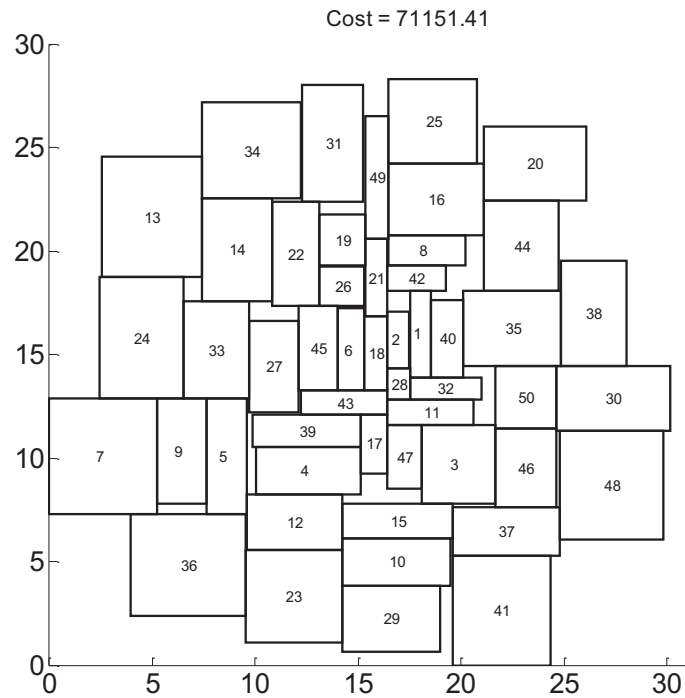


Figure 4: Optimal layout for problem OUH-050-EOS

Table II. Summary of Results

	Problem ID	Best Published		Best from PLANOPT	Proposed Algorithm
1	OUH-020-IMx	1,151.40	(McKendall, 2010)	1,157.00	1,140.00
2	OUH-028-MIx	8,640.00	(Mir, 1996)	6,447.25	6,289.64
3	OUH-050-EOS	71,291.40	(McKendall, 2010)	78,224.70	71,151.41

Conclusion

This paper presents a novel, efficient and effective algorithm for open-space, unequal area and hard module layout planning problems using a definition for Local Optimal Layout and a Near-optimality Hypothesis. This hybrid algorithm is a novel combination of an analytical technique (steepest descent) and a heuristic boundary search (corner search). The steepest descent has been used to determine the optimum direction of movement of a module. Corner search has been used at each placement of a module, where the modules are moved individually as well as collectively if these movements improve the objective function value.

The proposed algorithm brings significant, often dramatic, improvements over published techniques. The algorithm was tested on various well-known and large-scale

benchmark problems involving open-space, unequal area and hard module layout planning. For all these problems, the proposed algorithm furnished layouts whose cost surpassed all of the previously reported solutions including those obtained by the commercial software VIP-PLANOPT. In future, we plan to survey other published benchmark problems for open-space, unequal area and hard module layout planning for further testing of the efficacy of the proposed algorithm.

This research is expected to afford efficient procurement of superior layout outcomes and to facilitate cognitive, ergonomic, and economic efficiency of layout designers and operations managers. Applications of reported research are broad ranging including facilities layout design, VLSI circuit layout design, webpage layout design, newspaper layout design, cutting and packing dynamic memory allocations, multi-processors' scheduling, etc. It would also help improve the efficiency, efficacy, and scalability of layout optimization software and facilitate development of expert systems and decision support systems for relevant application domains. An ability to efficiently generate superior layout outcomes would also spur creativity of operations managers resulting in more insightful management of operations.

Appendix A

Let,	
NM	= Number of modules
SCM	= Set of candidate modules
PM	= Set of placed modules with their locations
TPM	= Total number of PMs
IM	= Incoming module
PMS	= Set of PMs with slopes in descending order
PMtemp	= Temporary set of PMs
Cost	= Cost of the objective function
CostPMtemp	= Temporary cost of the objective function
TPMO	= Total number of PMs with possible overlap(s)
PMO	= Set of PMs with possible overlap(s)
ϵ	= A small positive number

/Initialize/

Set Cost = ∞ and TPM = 0
 Randomly select IM \in SCM
 Place IM
 Add IM to PM
 TPM = TPM + 1

MainAlgo ()

For i = 2 to NM

Select another IM \in SCM
Call SearchBorder (
 Add IM to PM
 TPM = TPM + 1
 Calculate Cost
Call ImproveAnalytically(
For j = 1 to TPM
 Call SearchBorder (
 Call ImproveHeuristically(i)

SearchBorder ()

For k = 1 to TPM
 Place corners of IM or PM_j on that of PM_k
 Calculate Cost as newCost
 Save the minimum Cost in CostPMtemp
 Save layout in PMtemp
 If CostPMtemp < Cost
 Cost = CostPMtemp
 PM = PMtemp

ImproveAnalytically ()

For i = 1 to TPM
 Calculate S_i
 Sort PM modules by S_i & store in PMS
 For i = 1 to TPM
 Do While CostPMtemp < Cost
 Move PM_j \in PMS in the direction of S_i
 CheckOverlaps
 If TPMS $\neq \emptyset$
 RemoveOverlaps
 Calculate new Cost (CostPMtemp)
 If CostPMtemp < Cost
 PM = PMtemp
 Cost = CostPMtemp

ImproveHeuristically (k)

For i = 1 to k
 For d = 1 to 4
 Do While CostPMtemp < Cost
 Move PM_i in direction "d" by ϵ
 Save Cost in CostPMtemp
 Save layout in PMtemp
 If CostPMtemp < Cost
 PM = PMtemp
 Cost = CostPMtemp

References

- [1] A. Drira, H. Pierreval and S. Hajri-Gabouj, "Facility layout problems: A survey," *Annual Reviews in Control*, vol. 31, pp. 255-267, 2007.
- [2] A. Taghavi and A. Murat, "A heuristic procedure for the integrated facility layout design and flow assignment problem," *Computers and Industrial Engineering*, vol. In Press, 2011.
- [3] M. Anjos and F. Liers, "Global Approaches for Facility Layout and VLSI Floorplanning," in *Handbook of Semidefinite, Cone and Polynomial Optimization: Theory, Algorithms, Software and Applications*, 2010.
- [4] C.-L. Yang, S.-P. Chuang and T.-S. Hsu, "A genetic algorithm for dynamic facility planning in job shop manufacturing," *The International Journal of Advanced Manufacturing Technology*, vol. 52, no. 1-4, pp. 303-309, 2011.
- [5] M. Saif and M. Imam, "Computer Aided Layout Optimisation Model for Minimizing Traffic Volume," in *B.H.V. Topping, C.A. Mota Soares, (Editors), "Proceedings of the Fourth International Conference on Engineering Computational Technology"*, Stirlingshire, 2004.
- [6] S. Sahni and T. Gonzalez, "P-Complete approximation problems," *Journal of the ACM*, vol. 23, no. 3, pp. 555-565, July 1976.
- [7] S. P. Singh and R. K. Sharma, "A review of different approaches to the facility layout problems," *Int J Adv Manuf Technol*, vol. 30, pp. 425-433, 2006.
- [8] A. Ramkumar and S. J. N. Ponnambalam, "A new iterated fast local search heuristic for solving QAP formulation in facility layout design," *Robotics and Computer-Integrated Manufacturing*, vol. 25, pp. 620-629, 2009.
- [9] T. James, C. Rego and F. Glover, "A cooperative parallel tabu search algorithm for the quadratic assignment problem," *European Journal of Operational Research*, vol. 195, pp. 810-826, 2009.
- [10] A. Raoot and A. Rakshit, "A fuzzy heuristic for the quadratic assignment formulation to the facility layout problem," *International Journal of Production Research*, vol. 32, no. 3, pp. 563-581, 1994.
- [11] R. Ahuja and J. T. A. Orlin, "A greedy genetic algorithm for the quadratic assignment problem," *Computers and Operations Research*, vol. 27, p. 917-934, 2000.
- [12] A. Misevicius, "A modified simulated annealing algorithm for quadratic assignment problem," *Informatica*, vol. 14, no. 4, p. 497-514, 2003.
- [13] E. Talbi, O. Roux, C. Fonlupt and D. Robillard, "Parallel ant colonies for quadratic assignment problem. Future Generation," *Comput Syst*, vol. 17, p. 441-449, 2001.
- [14] K. Kado, "A investigation of genetic algorithms for facility layout problems," 1995.
- [15] Komarudin and K. Y. Wong, "Applying Ant System for solving Unequal Area Facility Layout Problems," *European Journal of Operational Research*, vol. 202, pp. 730-746, 2010.
- [16] A. R. McKendall Jr and A. Hakobyan, "Heuristics for the dynamic facility layout problem with unequal-area departments," *European Journal of Operational Research*, vol. 201, pp. 171-182, 2010.

- [17] D. Scholz, F. Jaehn and A. Junker, "Extensions to STaTS for practical applications of the facility layout problem," *European Journal of Operational Research*, vol. 204, pp. 463-472, 2010.
- [18] D. Scholz, A. Petrick and W. Domschke, "STaTS: A Slicing Tree and Tabu Search based heuristic for the unequal area facility layout problem," *European Journal of Operational Research*, vol. 197, pp. 166-178, 2009.
- [19] VIP-Planopt, "Engineering Optimization Software," 1 1 2010. [Online]. Available: <http://www.planopt.com/>. [Accessed 2 March 2011].
- [20] R. Sahin, K. Ertogral and O. Türkbey, "A simulated annealing heuristic for the dynamic layout problem with budget constraint," *Computers and Industrial Engineering*, vol. 59, pp. 308-313, 2010.
- [21] M. Dong, C. Wu and F. Hou, "Shortest path based simulated annealing algorithm for dynamic facility layout problem under dynamic business environment," *Expert Systems with Applications*, vol. 36, pp. 11221-11232, 2009.
- [22] B. H. Ulutas and A. A. Islier, "A clonal selection algorithm for dynamic facility layout problems," *Journal of Manufacturing Systems*, vol. 28, pp. 123-131, 2009.
- [23] M. H. Imam and M. Mir, "Cluster boundary search algorithm for building-block layout optimization," *Advanced in Engineering Software*, vol. 29, no. 2, pp. 165-173, 1998.
- [24] H. Samarghandi, P. Taabayan and F. F. Jahantigh, "A particle swarm optimization for the single row facility layout problem," *Computers and Industrial Engineering*, vol. 58, pp. 529-534, 2010.
- [25] M. F. Anjos and G. Yen, "Provably Near-Optimal Solutions for Very Large Single-Row Facility Layout Problems," *Optimization Methods and Software*, vol. 24, no. 4-5, August 2009.
- [26] M. Ficko, J. Balic, M. Brezocnik and I. Pahole, "Solving of Floor Layout Problem in Flexible Manufacturing System by Genetic Algorithms," *International Journal of Advanced Intelligence Paradigms*, vol. 2, no. 4, pp. 354 - 364, 2010.
- [27] M. Mir and M. H. Imam, "A hybrid optimization approach for layout design of unequal area facilities," *Computers and Industrial Engineering*, vol. 39, pp. 49-63, 2001.
- [28] M. H. Imam and M. Mir, "Nonlinear programming approach to automated topology optimization," *CAD Computer Aided Design*, vol. 21, no. 2, pp. 107-115, 1989.
- [29] M. Mir and M. H. Imam, "A gradient based method for module placement," *Computers and Electrical Engineering*, vol. 16, no. 2, pp. 109-113, 1990.
- [30] M. Mir and M. Imam, "Topology optimization of arbitrary-size blocks using a bivariate formulation," *Computer Aided Design*, vol. 24, no. 10, pp. 556-564, October 1992.
- [31] M. H. Imam and M. Mir, "Automated layout of facilities of unequal areas," *Computers & Industrial Engineering*, vol. 24, no. 3, pp. 355-366, 1993.
- [32] M. Mir and M. H. Imam, "Analytic annealing for macrocell placement problem," *Computers & Electrical Engineering*, vol. 22, no. 2, pp. 169-177, 1996.
- [33] Tam, "A simulated annealing algorithm for allocating space to manufacturing cells," *International Journal of Production Research*, vol. 30, pp. 63-87, 1992.

- [34] A. R. Ahmad, O. Basir, K. Hassanein and M. H. Imam, "An Effective Module Placement Strategy for Genetic Algorithms based Layout Design," *International Journal of Production Research*, vol. 44, no. 8, pp. 1545-1567, 2006.
- [35] A. R. Ahmad, *An Intelligent Expert System for Decision Analysis & Support in Multi-Attribute Layout Optimization*, University of Waterloo, 2005.
- [36] A. R. Ahmad, O. Basir, K. Hassanein and M. H. Imam, "A Hierarchical Placement Strategy for Generating Superior Layout Decision Alternatives," *International Journal of Operations and Quantitative Management*, vol. 11, no. 1, pp. 261-280, 2005.
- [37] P. Welgama and P. Gibson, "A construction algorithm for a machine layout problem with fixed pick-up and drop-off points," *International Journal of Production Research*, vol. 31, no. 11, pp. 2575-2590, 1993.
- [38] VIP-Planopt, "Engineering Optimization Software," 01 January 2010. [Online]. Available: <http://www.planopt.com/>. [Accessed 02 September 2012].
- [39] M. Mir and M. H. Imam, "Optimal Placement for Hierarchical VLSI Layout Design," *Microprocessing and Microprogramming*, vol. 25, pp. 177-182, 1989.
- [40] M. H. Imam and M. Mir, "Cluster boundary search algorithm for building-block layout optimization," *Advances in Engineering Software*, vol. 29, no. 2, pp. 165-173, 1998.
- [41] M. H. Imam and I. A. Tasadduq, "An extremely simple operation for drastic performance enhancement of genetic algorithms for engineering design optimization," *International Journal of Engineering Science and Technology*, vol. 2, no. 11, pp. 6630-6645, 2010.
- [42] I. A. Tasadduq, M. H. Imam and A. R. Ahmad, "Problem dimensions and best layouts," 04 September 2012. [Online]. Available: <http://pami.uwaterloo.ca/pub/rahim/TIA2012.xlsm>.
- [43] S. S. Heragu, *Facilities Design*, Lincoln: iUniverse, 2006.
- [44] J. A. Tompkins, J. A. White, Y. A. Bozer and J. Tanchoco, *Facilities Planning*, New York: John Wiley, 2002.
- [45] I. A. Tasadduq, M. H. Imam and A.-R. Ahmad, "A novel metasearch algorithm for facility layout problems," in *Proceedings of the 41st International Conference on Computers and Industrial Engineering*, Los Angeles, 2011.

ASB 2013
Antigonish, Nova Scotia

Abdul-Rahim Ahmad
Rowe School of Business
Dalhousie University, Canada

SOCIO-TECHNICAL PERSPECTIVES ON IMPEDIMENTS IN ADOPTION OF OPERATIONS RESEARCH & MANAGEMENT SCIENCE TOOLS

This paper explores ways of enhancing the efficacy of OR/MS techniques within a broader context. We examine differences and complementarities between a pure OR/MS and a Systems approach to problem solving. We attempt to identify some potent impediments in adoption of OR/MS tools from both technological and behavioral perspectives. Furthermore, we provide some recommendations for alleviating the inherent issues with OR/MS approaches.

Introduction

The disciplines of Operations Research and Management Sciences (OR/MS) have been defined in various ways by various researchers and organizations. One concise definition is: “[OR/MS are the disciplines] of applying advanced analytical methods to help make better decisions” (www.scienceofbetter.org). For over sixty years, OR/MS has been growing in complexity and usefulness. Indeed, OR/MS brought with itself new technologies and paradigms to aid effective decision-making (Taylor, 2013). More importantly, such decision-making paradigms and technologies have enhanced the effectiveness of interdisciplinary problem solving teams and added value and efficacy of OR/MS tools (INFORMS, 2013). The proliferation of information technology further facilitated the use of OR/MS tools by more people, rendering an increased capability to add value through OR/MS tools. This impact of information technology on OR/MS discipline is expressed by the way The Institute for Operations Research and the Management Sciences (INFORMS) defines OR/MS: “[OR/MS are] the professional disciplines that deal with the application of information technology for informed decision-making” (Ibid.).

Indeed, OR/MS disciplines are deemed responsible for substantial improvement in efficiency, efficacy, and productivity in industry, government, and research and development. Nevertheless, the traditional style of preaching and teaching OR/MS made many people perceive OR/MS as a panacea in the real world problem solving. In reality, some OR/MS practitioners may advocate it as the rational and scientific problem-solving regime. Conceivably, when people discover through hard experience that this discipline has its constraints in dealing with the majority of omnipresent socio-technical and ill-structured problems, a negative affect prevails. Nevertheless, the OR/MS tools in tandem with other soft decision-making and problem solving approaches could deliver a lot more than what some people perceive.

In this paper, our objective is to explore ways of enhancing the efficacy of OR/MS techniques within a broader context. We will also examine differences and complementarities

between a pure OR/MS and a Systems approach to problem solving. It is not our intention to cover these techniques in depth because of space and scope limitations. However, we have tried to identify some potent impediments in adoption of OR/MS tools from both technological and behavioral perspectives. Furthermore, our discussions are generalized in the sense that we would not be wading into any lists, or details, of OR/MS tools and techniques found in OR/MS texts.

The rest of the paper is organized as follows. We offer a brief historical perspective of the OR/MS discipline. Next, we discuss the value provided by OR/MS tools in various arenas. Subsequently, we delineate some core aspects of OR/MS tools and techniques followed by an outline of some of the problems with traditional OR/MS approaches. Next, we provide some recommendations for interdisciplinary and synergistic approaches needed in advancing applications of OR/MS as well as its adoption in more effective problem solving and decision-making. We conclude the paper with a summary and limitations of the paper as well as some future research directions.

Historical Perspective

We believe that a quick glance at the history and success of OR/MS as a discipline would enable putting some relevant arguments in a better perspective. The OR/MS has its beginnings in the Second World War, when British military authorities faced pressing problems that were not amenable to any known scientific methods (Pidd, 2001). These problems arose from new technologies and applications to organizing operations. In order to maximize utilization of military resources, especially radar installations, and to analyze the available data, military formed teams of scientists from a wide variety of backgrounds, aptly named as the Operational Research Group. The eventual success of the OR group in improving resource utilization and effective decision-making lead to the creation of similar group in the US military. In fact, it can be argued that the victory of allies in the Second World War is attributable to efficient and effective use of all kind of resources made possible through the application of OR (INFORMS).

In post-war scenario, British and US scientists realized that this approach of creating diverse teams for building and optimizing models could be used to great effect and benefit in business and the public sector (Pidd, 2004). The practice of employing OR/MS in the civilian domain soon spread throughout the developed world. By the end of the 1950s, OR/MS had established a set of problems and methods that it generally dealt with. Lists of such tools and applications can be found in any OR/MS text (Taha, 2010; Taylor, 2013). The spectacular rate of increase in computing power coupled with the development of more advanced modeling and optimization tools meant that more complex models and problems have been progressively tackled with OR/MS techniques.

The modeling/optimization approach remained a major constituent of OR/MS theory and practice throughout the 1960s and 70s. However, starting in the early 1980s, many OR/MS researchers and practitioners realized and acknowledged that the traditional OR/MS approach was not adequate in all situations (Ackoff 2001). Several researchers and practitioners expressed concerns about the ever-increasing mathematical bias and narrow focus on optimization in the

OR/MS. These people suggested that the systems thinking, the soft systems methodology, and other related techniques were more appropriate than mathematical programming for many complex, ill-defined, and difficult-to-model problems confronted by OR/MS practitioners (Ackoff 2001, Ahmad 2005). They extended the narrow quantitative optimization focus to incorporate thoughts from such disciplines as philosophy, sociology, and organizational behavior. Indeed, this multidisciplinary approach envisioned by proponents was in line with the earliest applications of OR during wartime when the OR Group consisted of experts from a wide array of disciplines and work domains.

The 1990s saw a revolution in computer and information technology that has had a significant effect on the OR/MS practice. During the recent decades, there was a dramatic proliferation of spreadsheets, easy access to inexpensive OR/MS software, simple communication and computational tools, the spread of OR/MS knowledge via the web, and other forms of modern information technologies. This led many more individuals getting involved in OR/MS applications, often with little OR/MS training or education or intellectual support. It has added to the burden of OR/MS to prove itself as a viable and value enhancing discipline when many pseudo-practitioners in OR/MS have only perfunctory knowledge of a small subset of OR/MS tools. The situation is aggravated because many such individuals cannot, and often do not know how to, effectively complement OR/MS tools with some other pertinent disciplines.

The Value of OR/MS

The OR/MS tools are ubiquitously creating value in organizations ranging from small to large multi-national corporations, in public sector to militaries, in terrestrial explorations to extra-terrestrial explorations, etc. Such applications of OR/MS tools are directed at achieving more with fewer resources. The availability of better and cheaper computing power provides more opportunities to create value and improve effectiveness through OR/MS. It helps provide quantitative and business insight into complex problems. Furthermore, it helps improve performance through a variety of modes, including analyses and recommendations about business processes, decreasing cost or investment, assessing the likely outcomes of decision alternatives and uncovering better alternatives, providing a superior basis for forecasting and planning, efficiently scheduling staff and equipment, improving productivity of organizations, increasing return on investment, improving quality, quantifying qualitative considerations, managing and reducing risk, achieving greater utilization from limited equipment, facilities, money, and personnel as well as supporting operational, tactical, and strategic planning.

Nevertheless, the field of OR/MS has expanded so much that most OR/MS analysts may only boast expertise in a small section of it. Consequently, they are highly likely to tailor/model any problem situation to fit in a small number of OR/MS techniques of their own expertise (Daellenbach and McNickle, 2012). Furthermore, one aim of OR/MS tools is to provide insights for informed decision-making (Ahmad *et al.*, 2008). Nevertheless, problem arises when users of OR/MS tools start taking the output of an OR/MS model as a decision rather than an input for decision support. In addition, there is a considerable gap between theory and practice of problem

solving techniques in the real world situations (Silver and Meal, 1998). There is a need to provide deep and rigorous treatment of OR/MS to practitioners without complicating mathematics (Ibid.).

The Traditional OR/MS Approaches

The traditional OR/MS approaches are analytic and reductionist (Jackson, 2000). A hierarchical decomposition approach is followed to explore how a system works leading to recommendations as to how things can be changed for better (Daellenbach and McNickle, 2012). It often consists of hard and technical ways to tame the problem, embracing scientific methodologies (Ibid.). An interesting analogy of OR/MS is that of “a doctor who, with little participation from the patient, indicates how health can be improved” (Keys 1991, pp. 2).

One simplifying tactic used by OR/MS practitioners is to focus on individual parts of the system and try to improve the performance of parts (Jackson, 2000). The assumption is that improving the performance of the parts of a system taken separately will improve the performance of the whole (Lawson, 2005). In reality, it may even destroy the whole system (Ackoff, 2001; Warfield and Staley, 1996). Nevertheless, dividing large projects into manageable parts has remained a viable problem solving strategy, which usually results in compartmentalization of information (Lawson, 2005).

From the early days of OR/MS, quantitative analyses were employed to provide some discernible and computable measures of effectiveness, enabling meaningful interpretations of the information generated. OR/MS provided useful information and insights to decision-makers for more informed and effective decision-making. Indeed scientific methods provide a basis for traditional OR/MS to relate theory to practice. However, OR/MS has traditionally assigned greater substance to its observance of scientific methods. In the next Section, we discuss some other important aspects of OR/MS as well as some problems associated with those.

Problems with the Traditional OR/MS Approaches

In OR/MS, the usual starting point is a relatively well-structured and tightly defined problem with clearly articulated performance measures and alternative courses of action. It proceeds with formulation of a model for the problem situation, collecting necessary data, and solving and validating the model. A tendency to follow a mathematically accurate approach to problem solving and perceiving OR/MS as exact science turns out to be an impediment in effective problem solving. Although such an exact approach is desirable in mathematical research, presuming it to be a necessary step in solving complex real world problems tends to mask the right perspective of the problem and proving a certain lemma becomes the focus of the exercise. Such an approach reinforces the view that the OR/MS academics largely solve academic problems quite removed from reality. This usual approach inherently contains several problems that may limit efficacy of OR/MS tools, some of which are discussed below.

Deterministic and Hard Optimization Paradigm

An optimization model consists of an objective function and a set of constraints. The goal is to find a solution that gives the best value objective function while satisfying all the constraints included in the model. However, often the solution obtained is the solution to the model and not to the actual problem (Lawson, 2005). Although some techniques support stochastic processes, OR/MS models and tools still deal with abstractions and simplifications of reality more appropriate for deterministic systems (Ackoff, 2001).

In a typical OR/MS approach, mathematical modeling precedes the generation of alternative courses. Consequently, there is often a futile and time-consuming exercise of formulating a mathematical model that might not even produce a feasible solution keeping in view the plethora of qualitative constraints in socio-technical systems (Foulds and Thanckekary, 2001). However, once an OR/MS analyst has spent a significant amount of time in formulating a complex, albeit rigidly and myopically defined, mathematical model, such cognitive biases as anchoring may hinder the search for a viable course. These cognitive biases, as discussed further subsequently, might impede the much-needed leaps of faith and the need for breaking away from the course already followed (Greenberg, 2010; Tversky and Kahneman, 1974). Resource limitations further aggravate the problem by providing a rationale for not looking further.

The hard optimization and mathematical programming approach particularly fails when addressing complex problems with conflicting objectives. It often provides a unique optimal solution against a single, simple criterion (Ahmad 2008; Triantaphyllou *et al.*, 1998). The solution is often based on a model that fails to capture many of the realities of the practical situation and includes only hard constraints (Ahmad, 2005). Furthermore, it often cannot provide answers within reasonable computational time (Foulds and Thanckekary, 2001). Even though OR/MS practitioners may come up with an optimal solution to a certain problem given the objective situation, some highly subjective considerations often constitute the imperative and decisive role in the selection of the solution alternative (Ibid.).

Problem Complexity

Another issue with traditional OR/MS approaches is the complexity of the problem. Complexity of a problem is often seen as the number of components, variables, and their combinations considered or needing to be considered in making a decision (Jackson, 2000; Warfield and Staley, 1996). It is commonly referred to as *combinatorial complexity* (Garey and Johnson, 1979). For instance, the scheduling of jobs on various machines available is a hard problem from combinatorial standpoint. The larger the number of machines or jobs at hand, the larger the number of possibilities that should be screened in search of an optimal solution. However, the real world systems also exhibit another subtle but critical sort of complexity – the *dynamic complexity* – discussed further subsequently. The dynamic complexity appears from complex interactions among various system variables and agents over time. In fact, many variables may change simultaneously, thus confounding the effects of any intervention into the system.

OR/MS still tends to focus on narrowly defined questions with more emphasis on breaking down problems into smaller and lower-level components that are more tractable to analysis and more immune to uncertainty (Drejer *et al.* 2000). The tradition of reductionism in OR/MS can be traced back to F.W. Taylor and his Scientific Management approach to industrial engineering. Although the current practice of OR/MS employs tools boasting a much higher level of scientific sophistication and efficacy than early scientific management techniques, Taylor's scientific management approach still provides a classic perspective to reductionism. Such reductionist approaches involved reducing a system to its simplest elements, analyzing them, and calculating how to improve each element – an antithesis of a much more plausible integrative approach discussed later in this paper. The underlying notion in scientific management is that observation and measurement provide a secure ground for analysis (Taylor, 2013).

Social Aspects in Socio-Technical Problems

The dynamic complexity may arise in seemingly very simple problem situations. For instance, most political/social problems seen from the so-called objective lens would appear as very simple problems that may be resolved by resorting to simple interventions. However, such interventions do not always help keep other variables constant and modeling the effects of variable of interest.

Furthermore, problems arising from social, behavioral, or political influences require an understanding of such influences before embarking on some OR/MS solution methodology. Due regard to social and behavioral determinants of operational success might provide very high returns without resorting to rigid OR/MS techniques for optimizing some tangible aspects of the system. For instance, studies in Sun Microsystems™ have shown how spending about \$20,000 in understanding and improving intangible determinants of product success yielded about \$150 million dollars return (Rhodes, 2000). An oft-cited classic example relates to the wait time at elevators in a building during the early days of OR/MS. Users of elevators constantly complained about the excessive perceived waiting time for elevators. Decision makers, and a series of OR/MS teams, embarked on a variety of futile interventions to reduce the waiting time of passengers. The solution finally came from an OR/MS professional who tried to break away from the discipline's usual approach of modeling a problem situation and then solving it optimally and tried to inspect the situation personally rather than relying on heaps of data available. The magic solution that followed the observation constituted of installing large mirrors in the elevator areas giving people a chance to appreciate theirs and others appearances discreetly while waiting (Taha, 2002). Interestingly, all complaints about excessive wait times suddenly vanished.

Another example of disregarding social and behavioral aspects of relates to Taylor's scientific management approach towards improving the efficiency of an individual and, purportedly, improving the workplace and business productivity. It disregarded many important dynamics of the whole system, such as potential compensation paid by businesses to workers suffering injuries in performing repetitive task in a rigid fashion. Union resistance and resulting social tensions were ignored as well as any other social aspect of socio-technical systems, as highlighted by the famous Hawthorne study (Greenberg, 2010). Consequently, even though the apparent efficiency of an individual worker during a short time period may be improved, the

overall business efficiency and productivity may be reduced and render the whole exercise of scientific micromanagement counter-productive.

Information Unavailability, Uncertainty, and Filtering

An OR/MS tool without data is like a person without soul. However, the data is often so scarce that it constrains meaningful application of quantitative and statistical tools available to us. Furthermore, the information we receive usually comes through various information filters and strong systematic assumptions (Sterman, 1994). In a socio-technical system, such filters may involve propaganda and control of information. In other technical systems, where data is based on measurements, these might involve distortions, errors, biases, uncertainty, ignorance, etc. (Ibid.). Indeed, measurement is based on selection – not only the selection of variables of interest but also the range, the format, the equipment, the environment, and the agent for measurements.

A classic example comes from the *Nimbus 7* project NASA launched in 1978, where the measurements of atmospheric ozone levels did not enable the detection of a gigantic ozone hole above Antarctica. It was only in 1985 when the first scientific evidence of this ozone hole was published, forcing NASA scientists to investigate the reasons of this failure. The cause was the computer program that was based on the belief that very low readings of ozone may only come from instrumental malfunctioning. Indeed, a revisit to the original data collected by *Nimbus 7*, before the filtering, showed that this ozone hole, with all its potentially disastrous consequences for our planet, could have been detected in 1978 (Meadows *et al.*, 1992; Sterman, 1994). In short, a measurement system was created that simply filtered important pertinent information.

Interestingly, such filtering is not only present in information procurement but also in the selection of the problem and its potential solution. The famous *garbage can model* illustrates how a problem is pseudo-randomly selected from various existing problems and then a pseudo-randomly selected solution approach is pursued (Greenberg, 2010). Indeed, solving the wrong problem, facetiously referred to as *type III error*, is not rare. Again, a classic example comes from the very early days of OR/MS discipline when scientists in the Royal Air Force's OR group started working on optimally fortifying fighter planes. The approach involved meticulous measurement and data collection regarding most hit or damaged parts of the warplanes returned from war missions. The underlying idea was to fortify the more vulnerable parts of the plane. The fallacy was that a wrong conception of vulnerability was employed. This fallacy was recognized when a senior air force officer pointed out that they were basing their analyses on those planes that have actually returned from combat missions.

Apart from scarcity and filtering of the problem-pertinent data, the available data often supports divergent and competing interpretations. In an attempt to draw inferences from the data, we must resort to systematic simplifying assumptions. Nevertheless, our mental and mathematical model formulations are constrained by what we resolved to define, observe, and measure and other 'externalities' are usually disregarded/unnoticed in problem solving and decision-making process (Tversky and Kahneman, 1974). Consequently, such information unavailability, filtering, and uncertainty often result in false pretenses as well as unintentional *anchoring bias*, which

refers to a cognitive bias that describes the common human tendency to rely too heavily on the first piece of information offered when making decisions (Greenberg, 2010).

Subjectivity and uncertainty in the domain knowledge has been classified by some researchers as incomplete, inconsistent, imprecise, and vague knowledge (Ahmad *et al.* 2005). The term *incompleteness* implies the unavailability of some of the information, necessitating the employment of heuristics and approximate reasoning. *Inconsistency* suggests the disparity or conflict in the knowledge, causing problems in using the available information. *Imprecision* refers to values that are loosely defined or measured inaccurately. *Vagueness* indicates the subjectivity in the estimate about some value or fact, causing difficulty in aptly understanding the information (Ahmad *et al.*, 2004).

Often, the preferences and constraints faced in solving problems are intrinsically incomplete, imprecise, inconsistent, and vague (Ahmad, 2004, 2005). In addition, the uncertainty tackled in operations research literature is largely limited to the uncertainty in the occurrence of a well-defined event such as uncertainty in demand forecasts. Nevertheless, many problems also involve uncertainty in the event itself due to incomplete, imprecise, inconsistent, and vague information. However, even under a relatively deterministic environment, the cost of procuring exact and complete information could often be prohibitive as well as implicitly involve systematic judgment (Ahmad *et al.*, 2004; Serman, 1994).

The limited, imperfect, and uncertain nature of information affects our decisions. The ambiguity arising from the confounded nature of interaction among system variables and multiple interpretations of linguistic terms significantly alters the human ability to identify the structure and parameters of a system from the examination of the system (Jackson, 2000; Warfield and Staley, 1996). Indeed, people often reason in terms of vague and context dependent concepts in dealing with situations where they encounter uncertainty (Turban and Aronson, 2001).

Such a multitude of tangible and intangible considerations in real world applications underscores the need for a composite/multi-criteria fitness evaluation regime combining various desirable characteristics (Triantaphyllou *et al.*, 1998). One reason is that different fitness criteria represent different perspectives and are very much likely to be competing and conflicting. However, the traditional approach in OR/MS to decision-making with multiple objectives is a fundamentally ad hoc methodology. Generally, it works with optimizing the most important objective, while meeting minimal aspiration levels on all other objectives. In short, arbitrary boundary choices are used with little justification for it, involving high degree of subjectivity. Such an ad hoc approach evades many fundamental issues in MCDM (Ibid.). A situation with more than one decision-maker further aggravates the problem.

Indeed, the uncertainty in the environment calls for flexibility in order to respond to changes quickly. However, most OR/MS tools are designed to achieve a rigidly defined optimality, undermining the very notion of flexibility and adaptability.

Dynamic Reality and Rationality

Another salient issue is the dynamic nature of reality and the problem situation. Indeed, decision-making is not a static process. It usually involves a planning horizon during which the environment encompassing the problem may change significantly. Irrespective of the degree to which we agree with such assertions, there is no denying the fact that most problem situations are dynamic in nature, as new and unintended dynamics may emerge through interactions among various components in a system or its ecology (Daellenbach and McNickle, 2012). Such a dynamic and evolutionary character is also true for requirements and preferences of clients, pronouncing the need for the OR/MS practitioners to adapt their problem-solving approach quite often. Indeed, such dynamism in preferences frequently results from client's interaction with existing or intermediate solutions, a phenomenon often referred to as *dynamic rationality* (Bouyssou and Vincke, 1998; Sterman, 1994).

Bounded Rationality

A decision strategy that is completely informed, perfectly logical, and oriented towards economic reward is referred to as *perfect rationality* (Simon, 1955). However, in practice, decision makers often resort to *bounded-rationality* reflecting an inadequacy resources (Greenberg, 2010; Simon, 1957). The most taxing issue OR/MS practitioners face is the *ill-structured* nature of problems that tend to be complex, novel, subjective, and uncertain (Ignizio & Cavalier, 1994). Such undertakings require a high degree of creativity and expertise (Ibid).

Beside all limitations, ambiguity, and complexity of information, humans may still interpret available information in ways that might be far from rational. The search for an objective measure for evaluation and comparison of alternatives adds to the constraints that hamper effectual decision-making (Simon, 1955, 1957). As such, decision makers often resort to *satisficing* rather than rational optimization (March and Simon, 1958). The term satisficing refers to the process of setting an adequate degree of acceptability for a solution to a problem and screening solutions till one surpassing the benchmark is found (Greenberg, 2010).

Cognitive Biases and Flaws

As mentioned, there are various biases involved in decision-making. For instance, decision-makers often resort to making presumptions regarding some aspects of the available information, an affinity referred to as *framing* error (Greenberg, 2010). Furthermore, even when objective information remains unchanged, people's judgment is strongly affected by the frame in which information is presented. In addition, cognitive biases for procuring and dealing with information in an error-prone fashion may emerge in unsuspecting ways. For instance, there is an instinctive human tendency to rely on more recent or readily available information, a phenomena commonly referred to as the *availability heuristic* (George and Jones, 2011).

Indeed, memory is distorted by hindsight, the availability and salience of examples, and the desirability of outcome. In addition, there is often a propensity to expose oneself only to the information that conforms to one's own analysis and belief of the situation, referred to as the

confirmation bias (Greenberg, 2010). Sometimes, such biases are a result of the *information overload* where more information is acquired or available than is necessary to make effective decisions (George and Jones, 2011). Research provides cogent evidence that even scientists and professionals suffer from many judgmental and cognitive biases (Kahneman *et al.*, 1982). Consequently, despite all virtues of an expert's judgment, judgment alone may not be an adequate assessment tool (Jackson, 2000; Lawson, 2005). Nevertheless, failure of foresight and judgmental flaws often make people blame the technology or tool for the failure.

Flaws in the Model Development Process

A mathematical model expresses the relationships among various components of the system in quantitative terms. Indeed, mathematical models are relatively easy to manipulate and facilitate exploration of the effects of changes in inputs (Daellenbach and McNickle, 2012). Nevertheless, a model should be developed with the purpose to solve the problem with a clear purpose, which is to solve the problem at hand (Churchman 1970, Jackson, 2000).

In order to address complex problems arising in industry, many OR/MS practitioners fall back on constructing complex models and try to solve them using optimization techniques (Ackoff, 2001; Foulds and Thanchekary, 2001; Sterman 2000). However, researchers have increasingly asked whether such an approach can deal successfully with many of the complex, ill-defined, difficult-to-model issues OR/MS practitioners are required to solve giving rise to other approaches and such soft systems methodologies as decision support systems, expert systems, soft computing, etc (Foulds and Thanchekary, 2001). Indeed, there are many difficult problems for which an exact/optimal solution approach would take forever, even with all the computing resources of the world at disposal of problem solvers, just to get a feasible or workable solution. However, many OR/MS practitioners assume that implementing sophisticated techniques will solve all problems (Ackoff, 2001). Furthermore, many OR/MS proponents stress the need for adopting a 'mathematically correct' solutions approach. In reality, many reasoning mechanisms not based on any sound mathematical basis have been very successful in practice (Karray and de Silva, 2004; Negnevitsky, 2011; Shafer, 1976; Taguchi, 1989; Turban and Aaronson, 2001).

Flaws in Model Testing and Validation

In OR/MS, model testing and validation is presented as essential to any problem solving efforts. It establishes the credibility of the model. Such efforts require both internal validity (checking that the model is logically and mathematically correct) as well as external validity (checking that the model actually a sufficient representation of reality). However, unfortunately, more efforts are expended to validate the model, to establish it to be the right model, than to test the model. This approach erodes the utility of the model and, more so, the credibility of the modeler and the OR/MS discipline. Model testing should instead be designed to uncover errors in the approach, understand limitations of the model, enhance the model efficacy, and seek a more appropriate model for effective decision-making.

This exercise should also be directed at testing the validity of assumptions underlying the model formulation, its robustness, and its sensitivity to various factors. Indeed, models may not

essentially be verified or validated as, by definition, all models are abstraction/simplification of reality (Jackson, 2000). Furthermore, a parsimonious model, one that explains a large part of the variability in the system and its behavior by employing the smallest number of system variables, is deemed a desirable outcome (Daellenbach and McNickle, 2012). Nevertheless, such desirability should not deter, or digress, the analyst from critically assessing the viability of the model and its underlying simplifying assumptions (Churchman 1970). Furthermore, the model should be tested for an appropriate time horizon as unrealistically short or long time horizon may distort the conclusion resulting from the whole exercise.

Erroneous Inferencing

Generally, the success or failure of the OR/MS methodology is gauged using performance metrics that have limited capability to capture the essence of success or failure. Once again, such performance metrics disregard the importance, sometimes even the existence, of organizational, behavioral, and political factors in success or failure of any solution methodology.

As mentioned, the multitude of tangible and intangible considerations in real world applications advocated the use of MCDM regime. Nevertheless, the subjectivity, uncertainty, and incommensurable units render MCDM paradigm inherently difficult to realize in many applications (Ahmad, 2008). In addition, the tendency of people relying on a pre-existing perspective of their skills in judging their performance, or potential for performing well, in present and future assignments results in overconfidence (Daellenbach and McNickle, 2012).

Opportunities for Increasing Efficacy of OR/MS Approaches

Despite an apparently long list of impediments in success and adoption of OR/MS tools, these tools have been quite successful in solving a variety of problems and this is expected to continue to hold in future (Pidd 2004). However, we believe that the OR/MS techniques can be more effectively employed when embedded within decision support and soft OR/MS paradigms. Nevertheless, most OR/MS texts fail to provide readers exposure to such approaches and confine themselves to computational details of quantitative techniques. However, it is interesting that the very first organized textbook on OR/MS clearly stressed the comprehensiveness of OR/MS as a systems approach (Churchman *et al.*, 1957). Unsurprisingly, the first two authors of that book were also the earliest and most vocal critics of the hard optimization approach that was fast becoming a norm in OR/MS at that time (Jackson, 2000).

We believe that teachings in OR/MS discipline should be complemented by its role in organizations providing some perspective of organizational dynamics and business realities. Fortunately, many academic programs such as the department of Management Sciences, University of Waterloo, realize this requirement and provide students with an essentially interdisciplinary outlook. In this Section, we discuss some interdisciplinary solution paradigms that offer promise to increase efficacy of OR/MS.

Systems Perspective

A systems perspective promotes the notion that the real world problems are part of a complex system and should be solved with an integrated perspective (Daellenbach and McNickle, 2012; Jackson, 2000). Such an integrated and holistic perspective provides the ability to appreciate that doing one thing does not essentially keep all other aspects of the system intact and that everything is linked to everything else. Different schools of thought in systems thinking seem to agree that a systems and integrated view of the world, though rare, is desirable (Doyle, 1997; Lane, 1994; Richardson, 1996). Just one decision or intervention may provoke reaction from others, the Counterintuitive Behavior of the social systems (Forrester, 1971; Daellenbach and McNickle, 2012). Even a single or simple change in one of the factors under consideration may result in unpredictable changes in other factors influencing the system (Doyle, 1997).

A systems approach typically starts with problem definition and proceeds with generation of alternatives, model formulation, and analysis/selection of alternative. Notably, model formulation follows generation of alternatives where we identify alternative courses, albeit vague ones, that satisfy the project objective. The organizational, behavioral, political, and environments constraints may dictate only a handful of different courses (Serman, 1994). Gathering information on such alternative courses helps in formulating mathematical models that accurately simulate the performance of each alternative.

The systems approach does not necessarily call for an analytical solution (Keys, 1991). A key principle is to scrutinize a problem from multiple perspectives and mull over long-term consequences of decisions, including behavioral, organizational, political, and moral implications. A systems perspective often requires interdisciplinary teams for problem solving. Indeed, it has been frequently argued by many researchers that the traditional rational/scientific thinking in OR/MS is largely based on reductionist and cause-and-effect modes and may not be able to cope with the complexity and ill-structured nature of most real-world problems (Ossenbruggen, 1994).

The systems approach also employs qualitative notions of synthesis and emergence that can be used to tackle ill-structured problems giving more weight to efficacy than validation. Such approaches lead to explanations of why phenomena are, rather than how those operate (Keys, 1991). Such an approach advocates both subjective evaluation of qualitative factors and an objective analysis of quantitative factors in decision-making. Nevertheless, a qualitative analysis often proves to be critical in problem solving. In essence, systems analysis is more of a process of discovery and learning than an analytical exercise in optimization (Ossenbruggen, 1994).

The systems approach is not as closely associated with science as tools used in the OR/MS approaches. The methodologies employed in a systems approach are rigorous but differ in fundamental ways from those of science. For instance, a systems approach may not instill the same reliance on quantitative/empirical data as does OR/MS. Furthermore, model validation and manipulation offer difficulties for systems approaches. However, these approaches provide ways to work when there are limits on effective use of OR/MS.

The relationship or interrelationship between the OR/MS and the systems thinking has been discussed at length by many researchers (e.g. Ackoff, 2001; Churchman 1957, Daellenbach and McNickle, 2012; Jackson, 2000; Jackson and Keys, 1985; Keys, 1991; Ossenbruggen, 1994; Silver and Meal, 1998; Sterman, 1994; Stainton, 1984). Some even argue that the OR/MS is only a strand in different version of the systems approach (Jackson, 2000).

Soft OR/MS Paradigm

As already mentioned, researchers have been advocating soft OR/MS approaches for several decades (Daellenbach and McNickle, 2012). They aptly indicate that an overwhelming majority of real-world problem scenarios violates many, if not most, of the assumptions underlying the hard OR/MS approaches. Indeed, behavioral, social, and political facets are largely ignored in hard OR/MS approaches. Soft OR/MS approaches were designed to tackle situations that are ill-structured, ill-defined, and involve different stakeholders with differing and possibly conflicting opinions. Some salient characteristics of such approaches include structuring the issues in problem situation instead of narrowly focused problem solving, facilitating dialogue among stakeholders and eliciting resolution of the problem from stakeholders themselves, instead of imposing a solution from the analyst, focusing the ‘what’ questions before taking up the ‘how’ questions, etc. The hard OR/MS deals mainly with the ‘content’ of the problem and gives little weight to the ‘process’. In contrast, soft OR/MS emphasize on both content and process starting with an attempt to obtain a reasonably comprehensive perspective and shared understanding of the pertinent issues (Karray and DeSilva, 2004).

There is a range of soft OR/MS methodologies starting from Game Theory to Theory of Constraints. A detailed discussion of these various soft OR/MS approaches is beyond the scope of this paper and inquisitive readers may find good reviews in literature (e.g. Daellenbach and McNickle, 2012; Jackson, 1991; Ossenbruggen, 1994). However, all such approaches involve active participation of stakeholders and, unlike hard OR/MS, these do not constitute some sequence of systems/phases whose output becomes the input to the next system. All these approaches are recognized as learning systems used iteratively and flexibly, drawing from philosophy, psychology, sociology, and organizational theory (Ossenbruggen, 1994). Several researchers have argued that the potential of various soft computing technologies is limited only by the imagination of their users (Negnevitsky, 2011; Ruan, 1997; Zha, 2003). However, there is no dearth of OR/MS practitioners who look down upon these soft approaches and advocate/practice less-pertinent hard optimization techniques even in highly subjective and uncertain problem domains (Ahmad, 2005). Nevertheless, OR/MS practitioners should realize that their proposed solutions would be implemented in wake of numerous organizational and

political considerations and soft thinking might provide effective decision courses. Disregarding such issues may greatly diminish the chances of their success.

Decision Support Paradigm

The Decision Support System (DSS) paradigm to problem-solving provides a means for assisting decision makers in retrieving, summarizing, and analyzing decision relevant data. Consequently, it results in a reduction in the cognitive overload faced by the decision maker(s). Research has shown that DSS techniques are useful in generating and evaluating a large number of alternative solutions and effectively helping decision-makers in arriving at better decisions (Turban and Aronson, 2001; Greenberg, 2010). However, it should be emphasized that a DSS neither automates the decision process nor imposes solutions. It simply provides analytical and information processing support in an interactive environment. Indeed, this soft decision-making paradigm has successfully been applied in a variety of complex and subjective task domains and power of the OR/MS tools can be enhanced by incorporating its models and methods within a decision support system (Ayyub, 2001; Foulds and Thanchekary, 2001; Marakas 2002; Turban and Aronson, 2001). The author's personal experience of successfully using Decision Support and Expert Systems paradigms and Soft Computing tools for solving highly complex, non-linear, subjective, and uncertain problem of layout optimization have sensitized him to the potential of such tools coupled with OR/MS methodologies (Ahmad, 2005).

Indeed, it is irrational to expect that a specific solution would surpass all others for every evaluation objective under consideration. Furthermore, the objectives could often be fuzzy and dynamic in nature. Consequently, the generation of superior alternatives in a flexible and automated manner through some decision support mechanism is critical to many planning processes. Decision Support Systems (DSS) represent a class of computerized information systems that utilize the knowledge about a specific application domain to assist decision makers by recommending appropriate actions and strategies (Ahmad *et al.*, 2008; Turban and Aronson, 2001). The DSS problem-solving paradigm provides a means for assisting decision makers in retrieving, summarizing, and analyzing decision relevant data. Consequently, it results in a reduction in the cognitive overload faced by decision makers. Research has shown that DSS techniques are useful in generating and evaluating a large number of alternative solutions and effectively helping decision-makers in arriving at better decisions (Ahmad, 2002; Turban and Aronson, 2001; Greenberg, 2010). However, it should be emphasized that a DSS neither automates the decision process nor imposes solutions. It simply provides analytical and information processing support in an interactive environment. Furthermore, a DSS should be much more than just a mere collection of models and solution techniques (Turban and Aronson, 2001). The essence of the DSS is the user-system interface that allows the planners to interactively conduct experiments, input decision rules and criteria, tackle issues with unstructured and semi-structured scenarios, and allow for multiple criteria decision-making as well as adapt soft constraints while providing flexibility, user friendliness, and convenience.

There is a need of developing decision systems that facilitate implementation of more sophisticated OR/MS tools giving people systems with which they can creatively solve their own problems (Silver and Meal, 1998). This need is more pronounced by the fact that the mantra in

the industrial and public sector organizations is to empower employees. OR/MS can provide them with the means to monitor, comprehend, intervene, and continuously improve their own performance. It is our belief that judicious institution of the decision support and expert systems paradigms may nicely bring together the scientific methods of OR/MS, the soft thinking of systems approach, and the learning from social/behavioral sciences to engender a more effective problem solving environment. We perceive these as complementary disciplines.

Simulation Modeling

Understanding the behavior of a system requires knowledge about how the state of the system changes over time. Collecting relevant data for understanding the behavior might not be possible because of various limitations such as high observation cost, long observation time, or even the non-existence of the system itself. In such scenarios, simulation or systematic imitation of the behavior of the system over time is a general approach and constitutes one of the most pertinent tools for an OR/MS analyst (Ahmad, 2005).

The typical absence of some encompassing, closed-form, and analytical fitness functions also render computer simulations a useful alternative (Ahmad, 2005). Such an approach would provide detailed analysis, modeling, and evaluation of complex problems (Bookbinder and Higginson, 1986). However, simulation models are not easily amenable to optimization and make procurement of a superior alternative difficult to achieve (Chan *et al.*, 1995).

Summary

We argued that the objective of OR/MS is not to provide 'optimal' solutions, as most traditional texts and practitioners profess, but to provide a basis for more informed and insightful decision-making. However, the usual reductionist approach adopted in many OR/MS techniques merely permits attention to details of each component of the system under study. In contrast, the systems thinking permits attention to the systematic and interactive roles of system components as an integrated whole. Often both modes of thinking are required to procure wider comprehension of a system. A soft OR/MS or a system thinking approach, though no panacea in efficacious problem solving, is a nice complement to traditional OR/MS. Its inherent approach of taking an integrated view and permissibility of soft/violable system boundaries is likely to lead to more encompassing, more insightful, and more effective decision-making.

At a meta-level, OR/MS, Systems approach, and Decision Support Paradigms may be conceived as sub-disciplines of a wider faculty. We believe that the limits of efficacious use of OR/MS approaches can be overcome by using systems approaches. Indeed, OR/MS approaches may be perfectly effective in some cases while in others systems approaches may be deployed with the same effect. Nevertheless, both approaches can make unique contributions towards effective decision-making in tandem with each other. We further believe that it is in the benefit of scientific community and the humanity that some effective meta-level framework guiding the effectual employment of both approaches be formulated. Such a framework would facilitate an

environment in which OR/MS (alternatively, Systems) disciplines have a positive view of relative strengths, an appreciation of the unique contributions of the other disciplines, and the potential learning from both to support effective decision-making and problem solving. Despite all arguments and counter-arguments in favor of one philosophy or the other regarding applicability of OR/MS, one aspect stands out: OR/MS practitioners should adopt an integrated systems methodology, whichever manner we define it, for solving real world problem plagued by behavioral, organizational, political, and environmental complexities, uncertainties, and constraints.

This paper is expected to further debate in this direction. However, this paper is limited in many ways, including a much-generalized discussion of tools, lacking any elaboration of such tools, due to by space and scope limitations. Furthermore, we have not expended much space on past efforts made in making OR/MS tools more effective and applicable by incorporating systems thinking in a wider problem solving approach. However, we believe, there is still a relative dearth in such efforts and some formulation of a wider framework for facilitating complementary utilization of OR/MS and other soft systems approaches in effective decision making.

References

1. Ackoff, R.L., "OR: After the post mortem", *System Dynamics Review*, Vol. 17, No.4, pp. 341-6, 2001.
2. Ahmad, A.R., Basir, O.A., Hassanein, K., "An Intelligent Expert Systems' Approach to Layout Decision Analysis and Support under Uncertainty", *Intelligent Decision Making: An AI-Based Approach*, *Studies in Computational Intelligence*, 97, (2008), pp 321-364.
3. Ahmad, A.R., *An Intelligent Expert System for Decision Analysis and Support in Multi-Attribute Layout Design*, Ph.D. Thesis, Systems Design Engineering, University of Waterloo, 2005.
4. Ahmad, A.R., "A Research Framework for Intelligent Decision Support in the Web Page Layout Design and Presentation Planning", M.S. Thesis (completed but not committed), Management Sciences, Faculty of Engineering, University of Waterloo, (2002).
5. Ahmad, A.R., Basir, O.A., Hassanein, K., "Decision Preferences, Constraints, and Evaluation Objectives in Layout Design: A Review of Modeling Techniques", *Proc. of the 5th Int'l Conf. on Operations and Quantitative Management (ICOQM-V)*, Oct. 2004, Seoul, Korea,.
6. Ayyub, B.M., *Elicitation of Expert Opinions for Uncertainty and Risks*, CRC Press, NY, 2001.
7. Bouyssou, D, and Vincke, P., "Preference Modeling", *Annals of Ops. Res.*, 80, (1998), pp. 147-180.
8. Churchman, C.W., "Operations Research as a Profession", *Management Sc.*, 17, (1970), pp. B37-53.
9. Churchman, C.W., Ackoff, R.L., Arnoff, E.L., *Introduction to Operations Research*, Wiley, 1957.
10. Daellenbach, H. and McNickle, D., "Management Science: Decision Making Through Systems Thinking", 2nd Ed., Palgrave Mcmillan, 2012.
11. Doyle, J., "The Cognitive Psychology of Systems Thinking", *System Dynamics Review*, 13(3), (1997), pp. 253-265.
12. Forrester, J.W., "The counterintuitive behavior of social systems", *Technology Reviews*, 73(3), (1971), pp. 52-68.
13. Foulds, L. and Thachenkary, C., "Empower to the People: How decision support systems and IT can aid the OR analyst", *OR/MS Today*, June 2001.
14. Garey, M.R., and D.S. Johnson, *Computers and Intractability*, W.H. Freeman Press, 1979.
15. George, G.M., and G.R. Jones, *Understanding and Managing the Organizational Behavior*, 6th ed., Prentice Hall, 2011.
16. Greenberg, J., *Behavior in Organizations*, 10th ed., Pearson Education, 2010.
17. Ignizio, J.P., T.M. Cavalier, *Linear Programming*, Prentice Hall, 1994.

18. INFORMS, Institute for Operations Research and the Management Sciences, <http://www.informs.org/>, 2013.
19. Jackson, M.C., *Systems Approaches to Management*, Kluwer, NY, 2000.
20. Jackson, M.C, Keys, P., "Introduction", *J. of Operations Research Society*, 36, (1985), pp. 753-755.
21. Kahneman, D., Slovi, P., and Tversky, A., *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1982.
22. Karray, F., De Silva, C., *Soft Computing and Intelligent Systems Design: Theory, Tools and Applications*, Addison-Wesley, 2004.
23. Lane, D., "With Little Efforts from our Friends: How System Dynamics and "Soft" O.R. can learn from each other", *System Dynamics Review*, Vol. 10, No. 2-3, pp. 101-134, 1994.
24. Lawson, D., *Engineering Disasters – Lessons to be learned*, ASME Press, NY, 2005.
25. Marakas, G.M., *Decision Support Systems*, Prentice Hall Inc., 2002.
26. March, J.G., and H.A. Simon, *Organizations*, Wiley, 1958.
27. Meadows, D.H., Meadows, D.L., Randers, J., *Beyond the Limits*, Chelsea Green Pub. Co., VT, 1992.
28. Negnevitsky, M., *Artificial Intelligence: A Guide to Intelligent Systems*, 3rd Ed., Pearson, 2011.
29. Norman, D. A., *Things That Make Us Smart: Defending Human Attributes in the Age of Machines*, Persius Books Group, 1993.
30. Ossenbruggen, P.J., *Fundamental Principles of Systems Analysis & Decision-Making*, Wiley, 1994.
31. Pidd, M., "The Futures of O.R.", *J. of Operational Research Society*, 52, (2001), pp. 1181-90.
32. Pidd, M., "The Futures of O.R.", *J. of Operational Research Society*, 55, (2004) pp. 791-800.
33. Rhodes, J.S., "Usability can save you money", <http://www.webword.com/moving/savecompany.html>, accessed on Nov. 02 2004, December 2000.
34. Richardson, G., *Modelling for Management*, Dartmouth Publishing Co. U.K., 1996.
35. Ruan, D., *Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks, Genetic Algorithms*, Kluwer, 1997.
36. Shafer, G., *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
37. Simon, H.A., "A Behavioral Model of Rational Choice", *Quarterly J. of Econ.*, 69, (1955), pp. 99-118.
38. Simon, H. A., *Administrative Behavior: a study of Decision Making Processes in Administrative Organizations*, 2nd ed., McMillan, 1957.
39. Stainton, R.S. "Applicable Systems Thinking", *European J. of Op Res*, 18, (1984), pp. 145-154.
40. Sterman, J.D., "Judgment under Uncertainty: Heuristics and Biases", *System Dynamics Review*, 10(2-3), (1994), pp. 291-330.
41. Taguchi, G., Elsayed, E., Hsiang, T., "Quality Engineering in Production Systems", McGrawHil, 1989.
42. Taha, H.A., *Operations Research: An Introduction*, 9th Ed., 2010.
43. Taylor, B.W., *Introduction to Management Science*, Pearson Ed., 11th Ed., 2013
44. Triantaphyllou, E., Shu, B., Nieto Sanchez, S., and Ray, T., "Multi-Criteria Decision Making: An Operations Research Approach", in J.D.Webster (Ed.) *Encyclopedia of Electrical and Electronics Engineering*, John Wiley, NY, CA, 1998.
45. Turban, E., and J.E. Aronson, *Decision Support Systems and Intelligent Systems*, Prentice Hall, 2005.
46. Tversky, A., and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases", *Science*, 185, (1974), pp. 1124-31.
47. Warfield, J.N. and Staley, S.M., "Structural Thinking: Organizing Complexity through Disciplined Activity", *Systems Research*, 13(1), (1996), pp. 47-67.
48. Zha, X.F., "Soft Computing Framework for Intelligent Human–Machine System Design, Simulation and Optimization", *Soft Computing*, 7, (2003), pp. 184-198.

REACTIONS OF POWERFUL CEOs TO ACCUSATIONS OF ILLEGAL ACTIVITY

We examine how CEOs use power to address public accusations of illegal activity by the firm. Results from a sample of S&P500 firms are consistent with models of cognitive framing and circulation of power, and suggest that powerful CEOs attempt to reduce uncertainty by resolving the accusation, either by accepting a conviction or an out-of-court settlement.

Introduction

Firms and executives that engage in illegal acts suffer long-term damage, in both performance (Baucus & Baucus, 1997), career longevity, and reputation (Beasley, 1996; Karpoff & Lott, 1993). Trevino and Youngblood (1990) explained illegal behavior in terms of individual “bad apples” acting in “bad barrels” of rationalized organizational and environmental factors (Anand, Ashforth, & Joshi, 2004; Ashforth, Gioia, Robinson, & Trevino, 2008).

The organizational perspective of “bad barrels” has contradictory results, with one study arguing that firms in munificent environments were less prone to illegal behavior (Staw & Szwejkowski, 1975), and another stating that they were more likely (Baucus & Near, 1991). This perspective also does not consider the role of individuals in wrongdoing, a key omission when considering legal repercussions.

At the individual level, characteristics that likely lead “bad apples” to commit illegal acts include misaligned incentives, age, experience, functional background, education, gender, charisma, and poor self-control (Zahra, Priem, & Rasheed, 2005; Zhang, Bartol, Smith, Pfarrer, & Khanin, 2008; Zona, Minoja, & Coad, 2013).

But we know less about how *suspected* “bad apples” respond to *allegations* of wrongdoing. The accusations may indeed have flagged illegal behavior, but on the other hand, executives may have done nothing wrong and the allegation may represent a retaliatory attack from a competitor, or a malicious accusation from a disgruntled former employee. The CEO is then faced with determining whether the allegation represents a threat or opportunity, and thus generates considerable response uncertainty, “the lack of knowledge of response options and/or an inability to predict the likely consequences of a response choice” (Milliken, 1987). Considering that CEOs have a unique and significant influence on firm actions (Jensen & Zajac, 2004), and a great impact on firm performance (Cannella & Hambrick, 1993), we expect CEOs to have a noticeable effect on how firms respond to allegations of illegal activity.

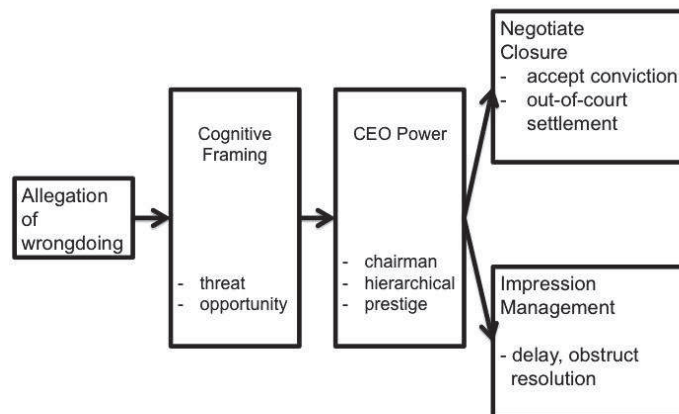
Wrongdoing can be seen as a highly risky activity (Mishina, Dykes, Block, & Pollock, 2010). However, with few exceptions, earlier studies that examined how executives take risks do not

discriminate between legal and illegal risk (Chattopadhyay, Glick, & Huber, 2001; Holmes Jr., Bromiley, Devers, Holcomb, & McGuire, 2011; Kahneman & Tversky, 1979). On the other hand, studies on executive wrongdoing tend to focus on individual characteristics, psychological effects, or financial incentives (Zahra et al., 2005; Zhang et al., 2008; Zona et al., 2013).

In this paper, we consider the power perspective (Pfeffer, 1992), which argues that CEOs dealing with risky situations require considerable power in order to secure the desired organizational response. We therefore ask when faced with an accusation of wrongdoing, how do powerful CEOs respond? Some CEOs may see a tactical opportunity and prefer to delay addressing the uncertainty by stalling and obstructing any investigation into the accusation in order to increase, defend or consolidate their power (Ocasio, 1994). However, others may insist on a rapid resolution of the allegation because it distracts resources and attention from achieving the strategic goals of the company (Ocasio, 1997), damages reputation (Hirshleifer, 1993), and because stakeholders tend to ignore the relative seriousness of illegalities, paints all wrong-doers with the same brush (Baucus & Baucus, 1997). They may also calculate that the financial punishment imposed by accepting a conviction may be insignificant compared to the reputational damage the allegation inflicts on both the firm and the executive (Karpoff & Lott, 1993; Lott, 1996).

We therefore combine theories of cognitive framing of risk (Chattopadhyay et al., 2001), and two principal theories of personal power -- agency theory (Jensen & Meckling, 1976), and the theories of circulation and institutionalization of power (Ocasio, 1994) -- to understand the rationales behind CEO responses to allegations of illegal activity. In short, we first argue that executives treat allegations as either a threat or opportunity. However, in order to act, they also require sufficient organizational power. We consider three forms of power: chairman, hierarchical, and prestige. Our model is shown in Figure 1. Our results suggest that powerful CEOs use their considerable hierarchical power to resolve allegations and reduce the response uncertainty, treating allegations as threats.

Table 1 Theoretical Model



Hypothesis Development

Cognitive Framing of Risk: Is the Allegation A Threat or Opportunity?

Allegations of wrongdoing represent a new pressure from the external environment. Cognitive framing theories argue that individuals make decisions depending on how they see their current situation relative to a reference target (Bromiley, Miller, & Rau, 2001). In short, do they consider the allegation as a threat or opportunity (Chattopadhyay et al., 2001)? One central model of cognitive framing, prospect theory (Kahneman & Tversky, 1979), suggests that individuals become risk seeking in the face of potential loss, since they underestimate the risk of additional potential loss and overweigh potential gains that could get them out of the loss position (Fiegenbaum & Thomas, 1988). Believing that the situation is undesirable, executives therefore shift attention towards acting in the external environment to respond to the allegation (Chattopadhyay et al., 2001). However, individuals are risk averse when facing potential gains, since they underweigh potential additional gains and overweigh risk of potential loss (Fiegenbaum & Thomas, 1988; Kahneman & Lovallo, 1993). They focus on action within the organization to address the allegation (Chattopadhyay et al., 2001) because they are satisfied that the current situation is better than their reference point. Research generally suggests that individuals in gain frames exhibit risk-averse behavior, while those in loss frames give rise to risk-seeking behavior (Holmes Jr. et al., 2011). In short, executives prefer risky, externally-directed responses if they feel they face loss, but they favor risk avoidance through internally-directed action if they believe they can gain from the allegation situation (George, Chattopadhyay, Sitkin, & Barden, 2006).

CEO Power

While the CEO may adopt a loss or gain frame in their decision, we also argue that they need sufficient power to act. Since the basis for CEO power is the ability to manage information uncertainty (Finkelstein, 1992; Hinings, Hickson, Pennings, & Schneck, 1974), our starting premise is that CEOs deal with accusations of illegal activity within their firms in order to defend, perpetuate or increase their own power through manipulation of information asymmetry. Executives wield at least two distinct forms of power: vertical power based on structure, and horizontal power based on social contacts outside the firm (Mitsuhashi & Greve, 2004), which we discuss next.

First, vertical or structural power provides CEOs with a greater ability to bend the organization to serve personal objectives (Pfeffer, 1992). From a resource dependence perspective, those with superior power manipulate values and meanings to perpetuate their power (Pfeffer & Salancik, 1978). CEOs with structural power are better able to control two key internal relationships within the firm: i) CEO-board of directors, and ii) CEO-subordinate employees.

Chairman Power. Vertical power occurs if the CEO simultaneously occupies the role of Chairman (*chairman power*) (Finkelstein & D'Aveni, 1994). Those who are responsible for overall management of the firm face a conflict of interest when, as Chair of the Board of Directors, they are also in a position to influence how to evaluate their own effectiveness (Zajac & Westphal, 1994). CEO-chairs can exert extra influence over board decisions through manipulating the board meeting's agenda, restricting available information, and steering discussions (Firstenberg & Malkiel, 1994). Agency theory (Jensen & Meckling, 1976) warns that CEOs with structural power may be better able to decouple their interest for personal gain from the interests of stakeholders. Powerful CEOs restrict a board's ability to effectively monitor and discipline the firm's managers by controlling the board agenda (Finkelstein & D'Aveni, 1994),

and steering board attention (Tuggle, Sirmon, Reutzel, & Bierman, 2010) towards the CEO's strategic priorities (Desai, Kroll, & Wright, 2003). They are therefore more likely to pursue action that either maximizes personal gain to the detriment of shareholders, or minimize personal employment risk (Boyd, 1995; Donohue, 2007). The emphasis on personal gain puts the CEO in a cognitive 'gain' frame, where they would favor risk avoidance through internally directed action. We therefore hypothesize that powerful CEOs will attempt to deflect, delay and obstruct any investigation to the allegation of illegal activity:

Hypothesis 1: There is a negative relationship between CEO chairman power and the likelihood that an accusation of illegal activity is resolved.

However, if the CEO focuses on minimizing employment risk, they adopt instead a 'loss' frame, and respond to the allegation with riskier, externally directed action. In this case, we hypothesize that powerful CEOs will favor resolution of the allegation:

Hypothesis 2: There is a positive relationship between CEO chairman power and the likelihood that an accusation of illegal activity is resolved.

Hierarchical Power. Even with the Board of Directors on board, the CEO must contend with subordinate employees' adherence to the plan. *CEO hierarchical power* refers to the top executive's formal position at the apex of the organization hierarchy that provides unique authority, resource control and network centrality (Astley & Sachdeva, 1984). The CEO's position induces hierarchical subordinates to obey not only because of the CEO's superior rank but also in the belief that the CEO has the right to exercise power (Astley & Sachdeva, 1984). Ocasio (1994) described two mechanisms to explain structural power: institutionalization and circulation. He argued that executives face two pressures that threaten their structural power – obsolescence and contestation – and that given enough power, they can deal with these pressures through power institutionalization. In this model, executives consolidate power by escalating commitment, institutionalizing rules and beliefs, and establishing networks of influence (Ocasio, 1994). This power attenuates employee resistance to change (Larsson & Finkelstein, 1999).

CEOs can therefore pursue these personal benefits and avoid stakeholder pressures by decoupling practices from policy (Westphal & Zajac, 2001; Zajac & Westphal, 1994). Even when the consequences of strategic action negatively affect the firm, powerful CEOs can escape personal sanction (Chatterjee & Harrison, 2003) through impression management (Elsbach & Sutton, 1992). CEOs with hierarchical power can therefore avoid shareholder pressures by decoupling action that maximizes personal gain from stakeholder pressure to act legitimately (Boyd, 1995). The emphasis on potential gain puts the CEO in a cognitive 'gain' frame, where they favor risk avoidance through internally directed action. Therefore:

Hypothesis 3: There is a negative relationship between CEO hierarchical power and the likelihood that an accusation of illegal activity is resolved.

On the other hand, CEOs recognize that the being associated with an illegal action may significantly jeopardize their personal careers, most likely obsolescence and ultimately, termination. The circulation of power model (Ocasio, 1994) argues that a CEO facing obsolescence will lead to a rise of challenges from other executives, who seek to contest the CEO's power. In short, facing allegations of illegal activity, powerful CEOs will fight back any challenges from within the organization by resolving the charge as quickly as possible. Because they adopt a 'loss' frame, we expect them to respond to the allegation with riskier, externally directed action. In other words, they favor resolving the allegation:

Hypothesis 4: There is a positive relationship between CEO hierarchical power and the likelihood that an accusation of illegal activity is resolved.

Prestige power. Horizontal, or prestige power, is formed through the ability of high-prestige actors to dominate lower-prestige actors because of their superior ability to access and control privileged information (Bunderson & Reagans, 2011). Prestige is therefore seen as a form of informal power (Finkelstein, 1992; Fredrickson, Hambrick, & Baumrin, 1988) based on associations with powerful and elite contacts as well as tacit knowledge and skills (Finkelstein, 1992), where executives control access to other social networks and the resources those networks control (Brockmann, Hoffman, Dawley, & Fornaciari, 2004). Prestigious executives have access to privileged information that can be used to reduce information asymmetry and aid decision-making (Carpenter & Westphal, 2001). These networks can supply managerial talent (Certo, 2003), or help executives stay up-to-date with the latest practices and procedures used in other firms (Davis, 1991).

Blau (1960) suggested that high prestige individuals enjoy a higher tolerance of deviant behavior than their lower-prestige counterparts. CEOs with prestige power are associated with more deviant strategies (Geletkanycz & Hambrick, 1997) and perform better (Geletkanycz & Boyd, 2011), even in the face of market or stakeholder disapproval (Chikh & Filbien, 2011; Haunschild, 1993) because they can block any potential sanctions (Perrow, 1961), they less likely to be sanctioned (Giordano, 1983), and they can more easily recover in case of failure (Chen, Hambrick, & Pollock, 2008). Because they adopt a ‘gain’ frame, we expect CEOs to prefer risk avoidance through internally directed action. Therefore:

Hypothesis 5: There is a negative relationship between CEO prestige power and the likelihood that an accusation of illegal activity is resolved

Method

Sample

We developed a sample of S&P500 firms between 2005 and 2010, where we identified 242 cases of illegal activity, including 106 cases of conviction or settlement, and 136 cases where a firm was only accused of wrongdoing. We selected S&P500 firms because they are prominent in the media, large and well capitalized, ultimately representing a benchmark and aspiration to other companies in the U.S. (Mishina et al., 2010). Illegal activities included: environment contamination, fraud, labour law violations, anti-competitive behavior, and misleading information. The sample spanned 52 2-dig SIC industries, with no single industry representing more than 9% of the sample and thereby suggesting that no single industry dominates. Data was drawn from COMPUSTAT/Research Insight, EXECUCOMP, BusinessWeek Executive Profiles, SEC 10K and DEF14A filings, and company websites.

Variables

Dependent variable. Because stakeholders tend to ignore the relative seriousness of illegalities, and only note whether or not some form of illegal action is present (Baucus & Baucus, 1997), we measured the incidence of illegal activity by noting whether or not at least one illegal activity had been associated with a firm during the calendar year. We noted cases of suspected of illegal activity (*Resolved*=0), and of resolution, either through conviction or out-of-court settlement as Resolved (*Resolved*=1).

We searched LexisNexis for evidence of actual or suspected criminal activity on 255 S&P500 companies between 2005 and 2010, selecting Newspapers, Newswires, US Newspapers & Wires, Non-US Newspapers & Wires, Major Newspapers, and Small Town Papers (US) as our sources. Through trial-and-error, we developed what we considered an exhaustive standardized keyword search script that was applied to each search in order to capture all possible cases of illegal activity. The majority of evidence was found in small-town news sources, while SEC violations were most prominent in wire sources. We cross-referenced each case across multiple news sources and found that there were no cases where sources contradicted each other. We found evidence of illegal activity in all but eight firms, suggesting that such activity (or allegations of such activity) is widespread across multiple industries. We summarize the distribution of types of illegal activity based on analysis of the representative texts from the news sources in Table 1.

Table 2: Types of Illegal Activity In The Sample

Type of Illegal Activity	Number of Cases	
	Accused	Resolved
Environmental contamination	9	18
Fraud	45	36
Labour Law Violation	29	10
Anti-Competitive Behaviour	34	20
Misleading Information	3	7
Patent Infringement	9	7
Other	7	7
Total	136	106

Independent variables. First, we captured *CEO chairman power* as a dichotomous (0/1) variable, indicating whether the CEO also occupied the Chairman role. Secondly, we used the CEO total compensation relative to the mean compensation of the other members of the top management team to capture *CEO hierarchical power* (Daily & Johnson, 1997; Finkelstein, 1992; Sanders, Davis-Blake, & Fredrickson, 1995). The total compensation was calculated as the sum of salary, bonus, and non-cash compensation (deferred income, stock grants, and stock options), with data obtained from EXECUCOMP and from SEC DEF14A proxy statements. Finally, we measured *CEO prestige power* by the number of the CEO's board directorships, where more directorships imply higher informal power (Finkelstein, 1992). The total number of corporate and non-profit boards was also measured. We also considered the relative prestige of the firm on whose board the CEO sits, since being a director of Google would be considered more prestigious than sitting on a board of a small, local company. We therefore noted the stock rating at the end of each calendar year of each company using the rating of the firm's general financial condition provided by Standard & Poor's Stock Quality Index (Finkelstein, 1992). The scale extends from a 0 for a non-rated firm, to 10 for a firm rated A+ by S&P. To determine the S&P board rating, we summed the scores of the firms on whose board the CEO sits. Principal component analysis confirmed a single *CEO prestige power* factor (eigenvalue = 1.89), explaining 63% of the total variance. Sampling adequacy (KMO = 0.63) exceeded the generally acceptable minimum of 0.5 (Hair, Anderson, Tatham, & Black, 1998). Reliability was acceptable (Cronbach α = 0.70) (Eisenhardt & Bourgeois, 1988; Finkelstein, 1992).

Control variables. To account for alternative explanations of responses to illegal activity, we included year, industry-level, firm-level, and CEO-level controls. Because some industries may be more prone to illegal activity than others, we included the two-digit SIC *industry*. At the firm

level, we considered *firm size*, *prior performance*, and *slack*. We controlled for *firm size* by measuring number of employees. Firm performance in the year prior to the investment accounts for the perception of potential gain or loss in risk attitudes of the executive based on prior firm performance instead of the accusation of wrongdoing (Wiseman & Gomez-Meija, 1998). High-performing executives are more likely to respond to allegations of wrongdoing because they felt additional “ratchet effect” pressure to continue improving performance (Freixas, Guesnerie, & Tirole, 1985). At the CEO level, we measured *CEO age*, *CEO tenure*, and *proportion of long-term compensation* relative to total compensation, all factors affecting the likelihood of illegal behavior (Daboub, Rasheed, Priem, & Gray, 1995; O'Connor, Priem, Coombs, & Gilley, 2006; Williams, Fadil, & Armstrong, 2005; Zahra et al., 2005; Zhang et al., 2008). We calculated the *proportion of long-term compensation* as (total compensation - total cash compensation)/(total compensation), where total compensation was the sum on the salary, bonus, deferred income, stock grants, and stock options from EXECUCOMP. Finally, we consider organizational slack since stakeholders tend to more closely scrutinize firms with less slack (McNamara, Halebian, & Dykes, 2008). *Available slack* estimates the resource cushion that is immediately available to be used to address the charge of wrongdoing, and is measured as the ratio of current assets to current liabilities (Latham & Braun, 2008). *Potential slack* captures the ability of a firm to secure resources with debt financing. Since some firms have zero debt, we used the debt-to-equity ratio to measure the inverse of potential slack (McNamara et al., 2008).

Estimation Model

We used a logistic regression to test the hypotheses given the dichotomous dependent variable. In essence, our model compares S&P500 firms that had outstanding accusations of wrongdoing to those who had resolved accusations either by out-of-court settlement or accepting a conviction. All independent variables were standardized with a mean of zero and a standard deviation of one to aid in interpretation of the influence of each variable, and to reduce the effects of multicollinearity (Jaccard & Turrisi, 2003).

Results

The variable statistics and correlations are listed in Table 2. The logistic regression coefficients, expressed as odds ratios, are listed in Table 3. Model I includes all control variables, while Model II adds the three hypothesized CEO power effects. A log likelihood test and a comparison of r^2 between the models suggest that CEO power has a significant effect on how CEOs choose to resolve allegations of illegal activity.

Hypothesis 1 predicted CEO-chairs are less likely to delay due to agency effects, while Hypothesis 2 argued that they are more likely. With no significant coefficient in Model II, neither hypothesis is supported. There is no evidence of an effect due to chairman power. Hypothesis 3 suggested that powerful CEOs avoid resolution of allegations as part of power institutionalization, while Hypothesis 4 predicted allegation resolution of allegations to avoid internal power circulation. A positive significant coefficient in Model II supports Hypothesis 4. Finally, Hypothesis 5 predicted prestige power afforded CEOs with the ability to decouple their fates from the fate of the firm, thus encouraging them to avoid resolution of the allegation. We found no evidence to support Hypothesis 5 in Model II.

Of the control variables offering alternative explanations of a CEO's response, we found support for prior performance and partial support for potential slack. CEOs of better-performing firms were much more likely to resolve the allegation, in line with the house-money effect of increased risk-taking based on playing with the recently gained money (Mishina, Pollock, & Bragaw, 2012;

Thaler & Johnson, 1990). This also provides evidence of their response to the “ratchet effect” pressure to continue improving performance (Freixas et al., 1985). We had mixed results concerning effects of resources on responsive action. Total firm resources and available slack had no measurable effect. However, firms with less potential slack (the inverse of our measure) were more likely to resolve the allegation, consistent with the idea that CEOs were sensitive to stakeholders more closely scrutinizing their firms (McNamara et al., 2008), perhaps because they feared too much slack would lead to less disciplined decision-making (Herold, Jayaraman, & Narayanaswamy, 2006).

Discussion

Our study examined two influences on how CEOs address accusations of wrongdoing. First, we considered cognitive framing: whether CEOs consider allegations as a chance for personal gain, or whether they see them as a threat of losing power. Secondly, whichever cognitive frame CEOs adopt, we assumed that CEOs required power to act. We examined three distinct forms of CEO power: chairman, hierarchical, and prestige. Our results only demonstrate an effect arising from hierarchical power, where CEOs were 3.5 times more likely to resolve an allegation through out-of-court settlement or conviction, than they were to leave the allegation open.

We found no evidence of chairman power or prestige power influencing how CEOs choose to address allegations of wrongdoing. This result is consistent with earlier studies suggesting that the effects on strategic action are different for the three sources of power. For instance, while structural power tends to increase with tenure, prestige power does not (Buchholtz & Ribbens, 1994). Likewise, structural power can help in recovery from bankruptcy, while prestige power does not (Brockmann *et al.*, 2004).

This paper naturally suffers from limitations. First, our sample represents data from an initial survey of secondary S&P500 data. Data from only 242 of the benchmark’s 500 firms has been collected. As well, our cross-sectional analysis does not consider longitudinal behavior of CEOs. In other words, serial illegal behavior by CEOs was not captured in our models. We also were not able to consider two possible combinations of firm actions and accusations, namely: i) where the firm has not been accused of wrongdoing, but in fact has done wrong, and ii) where it has not done wrong and they have not been accused of doing wrong. Finally, we only considered some rudimentary “bad apple” contexts in our models, including demographics and one psychological decision biases based on prior performance, but omitting more complex psychological attributes such as individual psychopathologies of criminal behavior (Babiak & Hare, 2006), charisma (Zona et al., 2013), and moral character (Eisenberg, 2000).

Table 3: Variable Statistics and Correlations

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11
1 Resolution	0.44	0.50	1										
2 Year 2006	0.08	0.28	-0.02	1									
3 Year 2007	0.12	0.33	0.09	-0.11	1								
4 Year 2008	0.16	0.36	0.01	-0.13 *	-0.16 **	1							
5 Year 2009	0.14	0.34	0.04	-0.12	-0.15 *	-0.17 **	1						
6 Year 2010	0.41	0.49	-0.08	-0.25 ***	-0.31 ***	-0.36 ***	-0.33 ***	1					
7 Industry	43.63	18.85	-0.08	-0.06	-0.04	0.06	0.06	-0.08	1				
8 Firm Size [^]	0.08	1.17	0.08	-0.05	0.02	-0.03	0.06	-0.03	0.06	1			
9 Debt/Equity [^]	0.04	0.52	0.10	0.09	-0.07	0.03	0.05	-0.08	0.07	0.05	1		
10 Available Slack [^]	-0.01	0.90	-0.02	-0.05	0.05	-0.12	0.05	0.05	-0.03	-0.22 ***	-0.38 ***	1	
11 CEO Age	54.72	6.30	0.07	-0.12	0.01	-0.12	0.02	0.14 *	-0.20 **	0.08	-0.08	0.01	1
12 CEO Tenure [^]	6.05	5.92	0.04	-0.06	-0.03	-0.03	0.03	0.07	-0.09	-0.10	-0.09	0.11	0.36***
13 Prior Performance	-0.01	0.08	0.09	0.02	-0.01	0.03	-0.00	0.01	-0.01	-0.09	-0.23 ***	0.24 ***	-0.01
14 Proportion Long-Term Compensation	0.80	0.19	-0.10	0.09	0.07	-0.12	-0.06	0.18 **	-0.07	-0.03	-0.11	0.01	0.07
15 CEO Power – Chairman	0.59	0.49	0.01	-0.03	-0.05	-0.06	0.04	0.08	-0.12	0.16	0.05	-0.06	0.30 ***
16 CEO Power – Hierarchical [^]	-0.00	0.39	0.13	-0.03	-0.04	-0.05	-0.01	0.12	-0.10	-0.05	-0.09	-0.12	0.21 **
17 CEO Power - Prestige	0.18	0.98	-0.01	-0.04	0.09	-0.01	-0.03	-0.01	-0.13	0.15	-0.05	-0.04	0.21 ***

Variable	12	13	14	15	16	17
12 CEO Tenure [^]	1					
13 Prior Performance	0.04	1				
14 Proportion Long-Term Compensation	0.11	0.14	1			
15 CEO Power – Chairman	0.35 ***	0.03	0.10	1		
16 CEO Power – Hierarchical [^]	0.09	0.02	0.39 ***	0.09	1	
17 CEO Power - Prestige	0.17 **	0.09	0.10	0.26 ***	0.07	1

[^]: variable is normalized through log transformation

*: p<0.05, ** p<0.01, ***: p<0.001

Table 4: Logistic Regressions of Likelihood of Resolving an Accusation of Illegal Activity (Odds Ratios)

Variable	Control Model I	CEO Power Effect Model II	Hypothesis
Intercept	3.08 (2.37)	7.06 (6.02) *	
Year 2006	0.92 (0.62)	1.04 (0.72)	
Year 2007	1.86 (1.13)	2.22 (1.41)	
Year 2008	1.06 (0.60)	1.13 (0.66)	
Year 2009	1.19 (0.69)	1.22 (0.72)	
Year 2010	0.88 (0.45)	0.91 (0.48)	
Industry	0.99 (0.01)	0.99 (0.01)	
Firm Size	1.17 (0.14)	1.26 (0.16) +	
Debt/Equity	1.76 (0.51) +	1.97 (0.59) *	
Available Slack	1.01 (0.17)	1.97 (0.59) *	
CEO Age	1.02 (0.03)	1.01 (0.03)	
CEO Tenure	1.12 (0.20)	1.20 (0.23)	
Prior Performance	46.67 (88.52) *	77.10 (150.27) *	
Proportion Long-Term Compensation	0.29 (0.22)	0.11 (0.10) *	
CEO Power – Chairman		0.83 (0.26)	H1 (-), H2 (+)
CEO Power – Hierarchical		3.45 (1.45) **	H3 (-), H4 (+)
CEO Power - Prestige		0.88 (0.13)	H5 (-)
N	242	242	
χ^2	16.95	27.90 *	
Pseudo r^2	0.05	0.08	
LR test compared to Model I		11.12 *	

+: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

Conclusion

In this paper, we asked when faced with an accusation of wrongdoing, how do powerful CEOs respond? We employed cognitive framing theory, agency theory, and the theories of circulation and institutionalization of power to understand the rationales behind CEO responses to public accusations of illegal activity by the firm. We compared three sources of CEO power and found that only one, hierarchical power, demonstrated any noticeable effect. Our results suggest that CEOs with hierarchical power were 3.5 times more likely to attempt to reduce uncertainty by resolving the accusation than leave the accusation unaddressed, either by accepting a conviction or an out-of-court settlement, suggesting that they treat allegations as a threat to their personal power. In short, powerful CEOs prefer to “take their lumps” now, protect themselves from emergent threats to their power, and move on with managing the company instead of delaying resolution of the accusation as a part of power institutionalization tactic.

References

- Anand, V., Ashforth, B. E., & Joshi, M. 2004. Business as usual: The acceptance and perception of corruption in organizations. *Academy of Management Executive*, 18(2): 39-53.
- Ashforth, B. E., Gioia, D. A., Robinson, S. L., & Trevino, L. K. 2008. Re-Viewing Organizational Corruption. *Academy of Management Review*, 33(3): 670-684.
- Astley, W. G. & Sachdeva, P. S. 1984. Structural sources of intraorganizational power: A theoretical synthesis. *Academy of Management Review*, 9(1): 104-113.
- Babiak, P. & Hare, R. D. 2006. *Snakes In Suite: When Psychopaths Go To Work*. New York: Harper Collins.
- Baucus, M. A. & Baucus, D. A. 1997. Paying the piper: An empirical examination of longer-term financial consequences of illegal corporate behavior. *Academy of Management Journal*, 40(1): 129-151.
- Baucus, M. S. & Near, J. P. 1991. Can Illegal Corporate Behavior Be Predicted? An Event History Analysis. *Academy of Management Journal*, 34(1): 9-36.
- Beasley, M. S. 1996. An empirical analysis of the relation between the board of director composition and financial statement fraud. *Accounting Review*, 71(4): 443-456.
- Blau, P. M. 1960. Patterns of Deviation in Work Groups. *Sociometry*, 23(3): 245-261.
- Boyd, B. K. 1995. CEO duality and firm performance: A contingency model. *Strategic Management Journal*, 16(4): 301-312.
- Brockmann, E. N., Hoffman, J. J., Dawley, D. D., & Fornaciari, C. J. 2004. The Impact of CEO Duality and Prestige on a Bankrupt Organization. *Journal of Managerial Issues*, 16(2): 178-196.
- Bromiley, P., Miller, K. D., & Rau, D. 2001. Risk in Strategic Management Research. In M. A. Hitt & R. E. Freeman & J. S. Harrison (Eds.), *The Blackwell Handbook of Strategic Management*: 259-288. Oxford, UK: Blackwell.
- Buchholtz, A. K. & Ribbens, B. A. 1994. Role of Chief Executive Officers in Takeover Resistance: Effects of CEO Incentives and Individual Characteristics. *Academy of Management Journal*, 37(3): 554-579.
- Bunderson, J. S. & Reagans, R. E. 2011. Power, Status, and Learning in Organizations. *Organization Science*, 22(5): 1182-1194.
- Cannella, A. A. J. & Hambrick, D. C. 1993. Effect of Executive Departures on the Performance of Acquired Firms. *Strategic Management Journal*, 14(Summer): 137-152.
- Carpenter, M. A. & Westphal, J. D. 2001. The strategic context of external network ties: Examining the impact of director appointments on board involvement in strategic decision making. *Academy of Management Journal*, 44(4): 639-660.

- Certo, S. T. 2003. Influencing Initial Public Offering Investors With Prestige: Signaling With Board Structures. *Academy of Management Review*, 28(3): 432-446.
- Chatterjee, S. & Harrison, J. S. 2003. Failed takeover attempts, corporate governance and refocusing. *Strategic Management Journal*, 24(1): 87-93.
- Chattopadhyay, P., Glick, W. H., & Huber, G. P. 2001. Organizational Actions in Response to Threats and Opportunities. *Academy of Management Journal*, 44(5): 937-955.
- Chen, G., Hambrick, D. C., & Pollock, T. G. 2008. Puttin' On The Ritz: Pre-IPO Enlistment of Prestigious Affiliates and Deadline-Induced Remediation. *Academy of Management Journal*, 51(5): 954-975.
- Chikh, S. & Filbien, J.-Y. 2011. Acquisitions and CEO Power: Evidence From French Networks. *Journal of Corporate Finance*, 17(5): 1221-1236.
- Daboub, A. J., Rasheed, A. M. A., Priem, R. L., & Gray, D. 1995. Top Management Team Characteristics and Corporate Illegal Activity. *Academy of Management Review*, 20(1): 138-170.
- Daily, C. M. & Johnson, J. L. 1997. Sources of CEO Power and Firm Financial Performance: A Longitudinal Assessment. *Journal of Management*, 23(2): 97-118.
- Davis, G. F. 1991. Agents Without Principles? The Spread of the Poison Pill Through the Intercorporate Network. *Administrative Science Quarterly*, 36(4): 583-613.
- Desai, A., Kroll, M., & Wright, P. 2003. CEO Duality, Board Monitoring, and Acquisition Performance; A Test of Competing Theories. *Journal of business Strategy*, 20(2): 137-156.
- Donoher, W. J. 2007. Employment Capital, Board Control, and the Problem of Misleading Disclosures. *Journal of Managerial Issues*, 19(3): 362-378.
- Eisenberg, N. 2000. Emotion, Regulation, and Moral Development. *Annual Review of Psychology*, 51(1): 665-697.
- Eisenhardt, K. M. & Bourgeois, L. J. I. 1988. Politics of Strategic Decision Making in High-Velocity Environments: Toward A Midrange Theory. *Academy of Management Journal*, 31(4): 737-770.
- Elsbach, K. D. & Sutton, R. I. 1992. Acquiring Organizational Legitimacy through Illegitimate Actions: A Marriage of Institutional and Impression Management Theories. *Academy of Management Journal*, 35(4): 699-738.
- Fiegenbaum, A. & Thomas, H. 1988. Attitudes Toward Risk and The Risk-Return Paradox: Prospect Theory Explanations. *Academy of Management Journal*, 31(1): 85-106.
- Finkelstein, S. 1992. Power in Top Management: Dimensions, Measurements, and Validation. *Academy of Management Journal*, 35(3): 505-538.
- Finkelstein, S. & D'Aveni, R. A. 1994. CEO duality as a double-edged sword: How boards of directors balance entrenchment avoidance and unity of command. *Academy of Management Journal*, 37(5): 1079-1108.

- Firstenberg, P. B. & Malkiel, B. G. 1994. The Twenty-First Century Boardroom: Who Will Be in Charge? *Sloan Management Review*, 36(1): 27-35.
- Fredrickson, J. W., Hambrick, D. C., & Baumrin, S. 1988. A Model of CEO Dismissal. *Academy of Management Review*, 13(2): 255-270.
- Freixas, X., Guesnerie, R., & Tirole, J. 1985. Planning under Incomplete Information and the Ratchet Effect. *Review of Economic Studies*, 52(2): 173-191.
- Geletkanycz, M. A. & Hambrick, D. C. 1997. The External Ties of Top Executives: Implications for Strategic Choice and Performance. *Administrative Science Quarterly*, 42(4): 654-681.
- Geletkanycz, M. A. & Boyd, B. K. 2011. CEO Outside Directorships and Firm Performance: A Reconciliation of Agency and Embeddedness Views. *Academy of Management Journal*, 54(2): 41-54.
- George, E., Chattopadhyay, P., Sitkin, S. B., & Barden, J. 2006. Cognitive Underpinnings of Institutional Persistence and Change: A Framing Perspective. *Academy of Management Review*, 31(2): 347-385.
- Giordano, P. C. 1983. Sanctioning the High-Status Deviant: An Attributional Analysis. *Social Psychology Quarterly*, 46(4): 329-342.
- Hair, J. F. J., Anderson, R. E., Tatham, R. L., & Black, W. C. 1998. *Multivariate Data Analysis* (5th ed.). Upper Saddle River, NJ: Pearson Prentice-Hall.
- Haunschild, P. R. 1993. Interorganizational imitation: The impact of interlocks on Corporate Acquisition Activity. *Administrative Science Quarterly*, 38(4): 564-592.
- Herold, D. M., Jayaraman, N., & Narayanaswamy, C. R. 2006. What is the Relationship Between Organizational Slack and Innovation? *Journal of Managerial Issues*, XVIII(3): 372-392.
- Hinings, C. R., Hickson, D. J., Pennings, J. M., & Schneck, R. E. 1974. Structural Conditions of Intraorganizational Power. *Administrative Science Quarterly*, 19(1): 22-44.
- Hirshleifer, D. 1993. Managerial Reputation and Corporate Investment Decisions. *Financial Management*, 22(2): 145-160.
- Holmes Jr., R. M., Bromiley, P., Devers, C. E., Holcomb, T. R., & McGuire, J. 2011. Management Theory Applications of Prospect Theory: Accomplishments, Challenges, and Opportunities. *Journal of Management*, 37(4): 1069-1107.
- Jaccard, J. & Turrisi, R. 2003. *Interaction Effects in Multiple Regression*. Thousand Oaks, California: Sage Publications.
- Jensen, M. C. & Meckling, W. H. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3(4): 305-360.
- Jensen, M. C. & Zajac, E. J. 2004. Corporate Elites and Corporate Strategy: How Demographic Preferences and Structural Position Shape the Scope of the Firm. *Strategic Management Journal*, 25(6): 507-524.

- Kahneman, D. & Tversky, A. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47: 263-291.
- Kahneman, D. & Lovallo, D. 1993. Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. *Management Science*, 39(1): 17-31.
- Karpoff, J. M. & Lott, J. R., Jr. 1993. The reputational penalty firms bear from committing criminal fraud. *Journal of Law and Economics*, 36(2): 757-.
- Larsson, R. & Finkelstein, S. 1999. Integrating Strategic, Organizational, and Human Resource Perspectives on Mergers and Acquisitions: A Case Survey of Synergy Realization. *Organization Science*, 10(1): 1-26.
- Latham, S. F. & Braun, M. R. 2008. The Performance Implications of Financial Slack During Economic Recession and Recovery: Observations from the Software Industry (2001-2003). *Journal of Managerial Issues*, XX(1): 30-50.
- Lott, J. R., Jr. 1996. The Level of Optimal Fines to Prevent Fraud When Reputations Exist and Penalty Clauses are Unenforceable. *Managerial and Decision Economics*, 17(4): 363-380.
- McNamara, G., Haleblan, J., & Dykes, B. J. 2008. The Performance Implications of Participating in an Acquisition Wave: Early Mover Advantages, Bandwagon Effects, and the Moderating Influence of Industry Characteristics and Acquirer Tactics. *Academy of Management Journal*, 51(1): 113-130.
- Milliken, F. J. 1987. Three Types of Perceived Uncertainty About the Environment: State, Effect and Response Uncertainty. *Academy of Management Review*, 12(1): 133-143.
- Mishina, Y., Dykes, B. J., Block, E. S., & Pollock, T. G. 2010. Why "Good" Firms Do Bad Things: The Effects of High Aspirations, High Expectations, and Prominence on the Incidence of Corporate Illegality. *Academy of Management Journal*, 53(4): 701-722.
- Mishina, Y., Pollock, T. G., & Bragaw, N. A. 2012. Know when to hold 'em, know when to fold 'em? Differentiating between loss aversion and the house money effect in CEO reactions to firm performance, *Academy of Management Conference*. Boston, MA.
- Mitsuhashi, H. & Greve, H. R. 2004. Powerful and Free: Intraorganizational Power and the Dynamics of Corporate Strategy. *Strategic Organization*, 2(2): 107-132.
- O'Connor, J. P., Priem, R. L., Coombs, J. H., & Gilley, K. M. 2006. Do CEO Stock Options Prevent or Promote Fraudulent Financial Reporting? *Academy of Management Journal*, 49(3): 483-500.
- Ocasio, W. 1994. Political dynamics and the circulation of power: CEO succession in U.S. industrial corporations, 1960-1990. *Administrative Science Quarterly*, 39: 384-416.
- Ocasio, W. 1997. Towards An Attention-Based View of the Firm. *Strategic Management Journal*, 18(Summer Special Issue): 187-206.
- Perrow, C. 1961. Organizational Prestige: Some Functions and Dysfunctions. *American Journal of Sociology*, 66(4): 335-341.

Pfeffer, J. & Salancik, F. R. 1978. *The External Control of Organizations*. New York, NY: Harper and Row.

Pfeffer, J. 1992. Understanding Power in Organizations. *California Management Review*, 34(2): 29-50.

Sanders, W. G., Davis-Blake, A., & Fredrickson, J. W. 1995. Prizes with strings attached: Determinants of the structure of CEO compensation. *Academy of Management Journal*(Best paper proceedings 1995): 266-270.

Staw, B. M. & Sz wajkowski, E. 1975. The Scarcity-Munificence Component of Organizational Environments and The Commission of Illegal Acts. *Administrative Science Quarterly*, 20(3): 345-354.

Thaler, R. H. & Johnson, E. J. 1990. Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science*, 36(6): 643-660.

Trevino, L. K. & Youngblood, S. A. 1990. Bad Apples in Bad Barrels: A Causal Analysis of Ethical Decision-Making Behavior. *Journal of Applied Psychology*, 75(4): 378-385.

Tuggle, C. S., Sirmon, D. G., Reutzell, C. R., & Bierman, L. 2010. Commanding board of director attention: investigating how organizational performance and CEO duality affect board members' attention to monitoring. *Strategic Management Journal*, 31(9): 946-968.

Westphal, J. D. & Zajac, E. J. 2001. Decoupling Policy from Practice: The Case of Stock Repurchase Programs. *Administrative Science Quarterly*, 46(2): 202-228.

Williams, R. J., Fadir, P. A., & Armstrong, R. W. 2005. Top Management Team Tenure and Corporate Illegal Activity: The Moderating Influence of Board Size. *Journal of Managerial Issues*, 17(4): 479-493.

Wiseman, R. M. & Gomez-Meija, L. R. 1998. A Behavioral Agency Model of Managerial Risk Taking. *Academy of Management Review*, 23(1): 133-153.

Zahra, S. A., Priem, R. L., & Rasheed, A. M. A. 2005. The Antecedents and Consequences of Top Management Fraud. *Journal of Management*, 31(6): 803-828.

Zajac, E. J. & Westphal, J. D. 1994. The Costs and Benefits of Managerial Incentives and Monitoring in Large U.S. Corporations: When is More not Better? *Strategic Management Journal*, 15(S1): 121-142.

Zhang, X., Bartol, K. M., Smith, K. G., Pfarrer, M. D., & Khanin, D. M. 2008. CEOs On The Edge: Earnings Manipulation and Stock-Based Incentive Misalignment. *Academy of Management Journal*, 51(2): 241-258.

Zona, F., Minoja, M., & Coad, V. 2013. Antecedents of Corporate Scandals: CEOs' Personal Traits, Stakeholders' Cohesion, Managerial Fraud, and Imbalanced Corporate Strategy. *Journal of Business Ethics*, 113(2): 265-283.

**NEW BRUNSWICK MODEL OF SHARED RISK PENSION PLAN:
HUMAN RESOURCE IMPLICATIONS IN AN ACADEMIC SETTING¹**

The province of New Brunswick has introduced changes to the Pension Benefit Act that facilitates the conversion of Defined Benefit Plans to a shared Risk Plan. This paper highlights the implications of one of the key provisions, career average pension, for recruitment retention and other aspects of human resource management in a university setting.

The declining interest earnings, increased mortality and the subprime mortgage triggered collapse of the financial markets globally have increased the solvency risk of a large number of Defined Benefit (DB) pension funds. The debt rating agency DBRS in its review of 451 pension plans in Canada and USA found that their deficit was almost four hundred billions of dollars². The problems of employer sponsored DB plans started attracting the attention of public at large due to some major corporate failures. For example, Nortel's pension plan had a deficit of around two billion dollars and had an impact on the retirement income of a large number of employees. In New Brunswick the failure of the paper mill at the Town of Nackawic almost devastated the community. As reported in the community Decision making tool kit- Nackawic Case Study "Pensioners lost 35% of their retirement funds in order to share what was left with all former workers. This situation contributed to further hostility between the citizens of the area - even resulting in rifts between fathers and sons, and between brothers who had all worked for the mill together."³ Recently, even publicly funded pension funds are facing deficit challenges. The increasing deficit in the public servants' pension plan started impacting the government's debt capacity and credit rating. The Province of New Brunswick set up a task force with mandate to

¹ I have benefited immensely from the discussions with the members of AEPP committee at the University of New Brunswick of which I am a member, plan actuaries and consultants and my colleagues at the Faculty of Business Administration UNB. My special thanks are due to Profs. Rob Austin, Vaughan Dickson, Lynn Haley, Robert Maher, Basu Sharma and Barbara Trenholm for their specific comments on an earlier version.. I am also thankful to the three anonymous reviewers for their constructive comments. .

² DBRS: Defined-benefit pension plans are DOOMED by The Canadian Press on Wednesday, August 22, 2012

³ <http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1238609114220&lang=eng>

review private pension plans and the mandate was later extended to publicly funded pension funds. Following the recommendations of the task force the government introduced changes to the Pension Benefits Act (1987) that paved the way for what is being called a “Shared Risk Pension Plan”(SRPP). - a hybrid between the defined benefit and defined contribution plan. The amendments to the Act and the accompanying regulations went into effect from July 1, 2012. Key features of the plan are available in the report produced by the Government of New Brunswick (2013). The New Brunswick model attracted a lot of both popular and professional media attention.[see for example Simpson (2012) and Wright (2012)] The two key elements of the SRPP are a career average pension accrual as against best average in a typical defined benefit plan and a contingent as against guaranteed adjustments for inflation.. The career average and its implication for human resource management is the focus of this paper and will be discussed in detail. Contingent indexing eliminated the guaranteed increase of pension for cost of living increases. Instead the plan focus on providing a funding level sufficient to meet a target of at least 75% of the increase over a period of twenty years. The actual increase in the pension to adjust for the inflation is contingent on the funding position. Any increase provided in a given year forms part of the base for subsequent years.

In addition to the two key elements affecting the benefits the SRPP provides for risk management tools and governance structure requirements. The risk management tools are aimed to achieve two goals.

The first one tries to maintain the funding policy such that the probability of the fund being unable to pay the base benefit is less than 2.5% over a period of twenty years .The second goal is to provide for pension increase equal to at least 75% of increase in the consumer price index during the pre-retirement period and to at least 75% of the promised increase in the current plan for the post retirement period on an average over a period of twenty years.

Over all the SRPP maintains the key element of a DB plan namely lifelong pension. However, the focus is providing sustainable base benefits and improving them based on the actual funding situation. A career average with or without enhancement is the key element to achieve sustainability and the focus of this paper is to identify any possible implications of the career average provision in SRPP on certain key areas of human resource management and industrial relations such as recruitment and retention, mid-career choices, salary structures and strike and lock outs in a university type of environment.

Retention

In a typical DB plan the pension accrual is based on the best five years’ salary. In other words every year of service will accrue a pension equal to x percentage of average salary of best five consecutive years. The key operative words here are consecutive years implying continuous service. In the career average, on the other hand, pension accrued is a percentage of the respective year’s salary. In an enhanced career average the accrued pension every year is increased to compensate for a targeted level of inflation. The career average or enhanced career average is a fundamental and significant change in the pension plan that has a major impact on the retention. The conventional DB plan is conducive for retention of employees. In such a plan continuous

service adds a lot to the accrued pension. An employee serving continuously for twenty years will accrue a larger pension than one who works for two ten years of service at different institutions even if the salaries were to be identical. If the accrual is say 2% of salary, twenty years of continuous service, will result in an accrual of 40% of average salary of years 16-20. On the other hand, someone with two ten years of service will accrue a pension of 20% of average salary of years 6-10 and another 20% of average salary of years 16-20. Given that the salaries are likely to be increasing over time in an academic setting, the pension for the first employee will be significantly larger. In the DB plan continuity of service has a high premium. An additional year of service increases the pension benefit in two ways:(1) By the value of the pension accrued for that year, and (2) By increasing the average salary and hence the pension accrued for past service. This premium on continuity is eliminated in the SRPP. Current year service or salary does not, under SRPP, have any impact on the pensions accrued in the past several years of service. The pension accrued over continuous twenty years will be the same as that over two ten year periods if the salaries are identical. In other words pension does not act as a tool of retention. In the DB moving or relocating has a negative impact on pension accrual if the employee does not move to another institution where the pension is transferable. SRPP on the other hand has eliminated the need for portability completely.

The pension deterrent being neutralized, one may expect more relocation from New Brunswick universities than before. I conjecture that relocation will occur in two categories of academics. Apart from monetary aspect of pension and salary, job security provided by tenure is an important factor in relocation for academics. Among tenured faculty members those who could get a tenured job in another institution will be the ones more likely to move than ones that need to work towards tenure in another institution. It is also well known that universities are reluctant to grant tenure on appointment unless the candidate's performance has been extremely good. Highly achieving recently tenured faculty members may therefore be the prime candidates to move to Ontario and other western provinces (or for that matter other eastern provinces) where comparable salaries are higher. In fact according to the CAUT almanac the 2010-11 average salary of assistant professor in New Brunswick was \$78461 as against the Canadian average of \$89681. Manitoba was the only province that had a lower average. These young academics are the ones who have overcome the deterrent of job security by their performance and the pension deterrent has been eliminated by SRPP. Thus, an unintended consequence may be that New Brunswick is likely to lose highly qualified young faculty members to other provinces. For employees at the middle level the mobility is perhaps very limited as host institutions in general prefer not to hire a tenured faculty at the professorial level unless the appointment is for a funded chair or for an administrative position.

The other group that may be motivated to relocate consists of faculty members that are a few years away from their planned retirement. Among this group those who have been performing well may have the ability to move but not with tenure. However, if they have only a few years prior to their planned retirement they can secure non tenured contractual positions for a few years. These are individuals who continue to be performing at a high level, mentor junior faculties, have research grants and graduate students. In the traditional DB scheme it is disadvantageous to leave towards the end of the career as the average salary will affect the pension accrued in prior years. In a SRPP conversion the average is frozen at the date of conversion and salary beyond that time

will not impact the average. Those who were planning to move out of province post retirement due to family connections or other cultural reasons may accelerate their move if they can secure a term position. Anecdotal evidence indicate that moving towards Montreal, Ottawa, Greater Toronto and Vancouver after retirement is common among first generation immigrant academics due to the size of the cities and prevalence of sizable ethnic population in these localities making ethno cultural activities viable at the social level and the availability of consulting opportunities at the professional level.

Another key provision in the SRPP that impact retention is that the contingent indexing is applied across board to the actives and the retirees from the effective date. By not grandfathering the index provision there is no special incentive to retire just prior to the implementation of SRPP. If grandfathering has been provided then pension consideration might have accelerated the decision of a large number of members that are closer to normal retirement age for pension purposes to retire prior to SRPP in order to have guaranteed indexation. Ideally a university would like a selective group of senior underachieving members retire as it would give the university an opportunity to renew by bringing in new young faculty members who can be mentored by active senior members. The absence of a grandfathering provision under SRPP prevents an exodus of senior members, but the career average provision facilitates the mobility of high achieving senior members..

Overall SRPP will result in an adverse selection problem for New Brunswick universities with DB plans. SRPP will motivate young high achievers to relocate, accelerate the retirement of senior high achieving members and will not provide any incentive to retire for other senior members that lack the mobility to relocate. The ease of mobility will have some other effects at a macro level. The ease of mobility reduces the transaction cost in the system and thus adds to efficiency. In the earlier example the host institutions can attract the talent that is relocating at a lower cost than before. The ease of mobility will also make certain organizational change processes easy as those who are resistant to changes can now move at a lower cost. The above discussions lead to the following proposition.

Proposition 1

Given other things constant the turnover of academic faculty in the SRP regime is expected to be higher than that under a DB regime. Further the higher turnover will be pronounced in two stages of the career: years immediately following a successful tenure and a few years prior to normal retirement.

Recruitment

As alluded to earlier the average salaries of New Brunswick universities are not competitive. However, in the total compensation pension benefit is one of the key elements. The benefits and the contribution rate by employer and employees of Ontario Universities are provided in Table 1 and 2. The DB plans of New Brunswick universities are comparable in structure. However, one cannot say the same about the contribution rates. In general, the employer contribution is more than the employee contributions in Ontario universities. Universities in Alberta and the University of New Brunswick are the only plans in the university setting with fifty / fifty contribution rates.

[Place Table 1 and Table 2 here]

The DB plans that are in deficit will have an amortized past service cost which increases the contribution required by the new employee for the benefit promised. In a conversion to SRPP from a DB the past deficits are to be eliminated and funded ratio is brought to be in excess of 100%. This process adds additional burden to new recruits. The conditional indexing and career averaging reduces a substantial portion of liability. In spite of reduction in benefits the contribution is not expected to decrease due to conservative policies needed to ensure the viability of the plan. As the government report itself indicates, "For new employees and early career employees, there is greater certainty that their pension will be there when they need to collect it, though they will likely pay a marginally higher contribution" (see page 13 of the report). One can define the value of the pension plan to a member as the difference between the total (employer and employee) current service cost of benefits and the contribution of the member. As the current service cost represents the present value of the expected benefits, the excess of that over member contribution represents the value of benefit received from employer. By the very design SRPP in order to meet the risk management goal, requires a contribution in excess of the current service cost. If pension benefit has to be an attractive tool for recruitment the value to member has to be maximized. However, in a conversion to SRPP from DB career average provision and the contingent indexing reduces the value of the benefit to the employees. If the plan is running a deficit at the time of conversion the benefit reduction reduces the deficit by this contribution in kind from the employees and retirees with no matching contribution from the employer. In other words the employer gains from the elimination of the unamortized past service cost and the burden is entirely shifted to the active members and the retirees of the plan. The only way to increase the value of member benefit is for the employee to contribute more at least for a temporary period or provide upfront lump sum contributions to offset the unamortized past service cost obligations. This leads to the next proposition.

Proposition 2

Given other things constant adoption of SRPP will make universities in New Brunswick less competitive

Strike and Lockouts

The DB plans place emphasis on continuous service for pension calculations. Hence, a period of strike or lockout may amount to interruption of service. It is customary to have in the back to work protocol provisions to negate explicitly the discontinuity effect. Thus, in a DB situation cost of strike (lockout) extends beyond the loss of pay during that time to major reduction in future pensions. The employer, therefore, has a higher bargaining power as the employees need them to yield on the continuous service treatment. Under SRPP regime the continuity of service has no impact on the pension. The impact of continuity of service on other benefits like vacation pay etc., though important, do not have the level of impact as continuity of service has on pension earnings in a DB situation. Thus, SRPP has removed one major deterrent for strike and reduced the cost of lockout significantly. The SRPP, in my view, thus tilts the balance of power to employees.

Mid-Career Temporary Administrative Positions

In the university setting a significant amount of administrative positions such as department heads and assistant deans are filled by academics and are compensated by stipends that are not part of their base salary. Once they leave the position the stipends will not be paid. In the typical DB plan members that accept such position in the middle of the career will be contributing to the pension plan based on the extra stipends. However, that extra contribution may not have any impact on the pension they may accumulate at retirement. This is because the final average is likely to be more than mid-career salary with stipend. In other words, two individuals with identical salary structure - one with mid-career stipend and another without - will receive the same pension at retirement. Under SRPP, on the other hand, pensionable salary at any point in time of the career will count towards final pension. Given other things constant I conjecture that under SRPP there will be additional motivation for accepting administrative positions even in the early and mid-career.

Salary Structure

The salary structure of academic institutions like universities is typically based on a career progression or seniority based structures. In such a scheme at the entry level the salaries are low and for every year of service a salary increase is provided other than the competitiveness adjustment and or inflation adjustment. In other words instead of working on the average salary throughout the career only adjusted for inflation, academics typically start with below average pay and end with above average pay towards the end of the career. The time based increase is typically called a Progress Through Rank (PTR) adjustment. In a DB Plan the best years' average incorporates all the past PTR adjustments and the inflation adjustments on them. In a SRPP regime it is important that the PTR adjustments are included in the computation of the base benefit. If PTR adjustments are to be treated as ancillary benefits, then only a target percentage of PTR, contingent on the funding level, may enter the base pension that is subject to inflation protection. In such a scenario, the value of the pension benefit to the members will be significantly different from a typical DB plan. Since PTR adjustments are the same dollar amount across ranks, the impact will be more for the junior members because the PTR adjustment, as a proportion of their salary, will be much higher than that for a senior member. In universities where pension and wage negotiations are separate pension negotiators have to be aware of the impact of PTR adjustments and its impact on the pension accrual.

Contract Academic Staff

The SRPP regime provides opportunity for a contract academic staff who teaches a course or two in a year to be included in the pension scheme. In the DB plan, time of service in terms of years and continuity of service to calculate the best average pay are important. The SRPP accrual is based purely on salary earned in a year and whether the salary is earned through a continuous employment or not is irrelevant. The benefits are also based on salary and hence the time is not a factor. Universities over years have been using more and more contract academic staff. Unlike medical and other disability benefits that have complications in terms of the need for

determination of time period when one is employed, for how long the coverage should last and the like the pension benefit is very easy to compute. Therefore, one may expect that there will be demand from bargaining units that represent contract academics, graduate teaching assistants and research assistants to demand pension benefits. Administratively though it will increase the number of members with small contribution and benefits thus increasing the cost of administering pension plans. A defined contribution plan may be an efficient alternative in such situations.

Voluntary Contributions

Even though it is not contemplated currently SRPP regime is easily amenable for added voluntary contribution from members to enhance her pension. This is because the benefit is easily related to the contribution. If, for example, the pension accrual is $x\%$ of salary and the total contribution by both parties put together is $y\%$ of salary then an extra contribution of $\$z$ will result in an accrual of $\$(z \times x/y)$. This is similar to situation where additional contribution is made due to administrative stipends except that there will not be an employer contribution. Along the similar lines members that have left the plan but have not taken out the vested amount may be in a position to make voluntary contributions. These are not part of the current SRPP, but conceivable to be the future additions. The constraints for such developments may be from the ability to have the group profile reasonably intact so that actuarial assumptions that determine the rates are not fundamentally altered.

Summary

In this paper I have argued that the SRPP regime is conducive for young academics to move to other provinces as it does not place any premium on continuous service. Further senior academics may accelerate their decision to retire and move to desired locations if they are capable of finding even non tenurable contract jobs. Furthermore, if universities convert to SRPP from a deficit DB plan the value of the pension benefit is likely to be very low for new recruits. Unless the universities are willing to provide additional unilateral funding upfront to the pension, as done by City of Fredericton that adopted SRPP, or plan for a temporary increase in its contribution for a number of years, they will lose their competitiveness in the market for academics. The problem is exacerbated by the fact the salary structure in New Brunswick for academics ranks is almost at the bottom among provinces. Of course this analysis is based on the effect of career average provision alone. The risk management aspects may mitigate some of these effects. I have also indicated possible future developments in terms of inclusion of contract academic staff and possibility of voluntary contributions. All of these issues will add new challenges to the human resource managers of the universities in New Brunswick and other jurisdictions that adopt the SRPP principles.

Table 1. Comparison of Ontario DB Plan Benefits

DB Pension Plan	Average Earnings (yrs.)	Benefit Rate After Age 65	Benefit Rate Before Age 65	Subsidized Payment Form	Automatic Indexing (% of CPI)	Eligibility for Unreduced Early Retirement
University of Toronto						
Faculty	3	1.5% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	75%	age 60 + 10 years
Staff	3	1.6% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	75%	age 60 + 80 points
University of Ottawa	5	1.3% / 2.0%	same	With Spouse: 60% J&S with G5 Without Spouse: LG5	CPI < 3%: 100% up to 2% CPI ≥ 3%: CPI – 1 %	age 60, or 90 points
McMaster University (Salaried Plan)	4	1.4% / 2.0%	same	With Spouse: 50% J&S Without Spouse: LG7	excess interest (threshold at 4.5%)	80 or 85 points by staff group
University of Waterloo	3	1.4% / 2.0%	same	LG10	100%	age 62
Ryerson University	5	1.35% / 2.0%	2.0%	With Spouse: 60% J&S Without Spouse: LG10	100%	90 points, or age 60 + 20 years
University of Guelph						
Professional Plan	3	1.5% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	CPI – 2%	age 55 + 85 points
Staff Plan	3	1.6% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	CPI – 2%	age 55 + 85 points
University of Windsor (Staff Plan)	5	1.5% / 2.0%	same	LG5	50% of excess interest (threshold at 6%) up to 50% of CPI	age 65

DB Pension Plan	Average Earnings (yrs.)	Benefit Rate After Age 65	Benefit Rate Before Age 65	Subsidized Payment Form	Automatic Indexing (% of CPI)	Eligibility for Unreduced Early Retirement
Trent University						
Faculty Plan	3	2.0%	same	With Spouse: 60% J&S Without Spouse: LG10	excess interest (threshold at 6.5%) up to 50% of CPI	age 65
Staff Plan	3	2.0%	same	With Spouse: 60% J&S Without Spouse: LG10	excess interest (threshold at 6.5%) up to 50% of CPI	age 60 + 80 points
OTPP	5	1.55% / 2.0%	2.0%	With Spouse: 50% J&S Without Spouse: LG10	100% (pre-2010 benefits); 50% (post-2009 benefits) plus top-up to 100% based on plan's funded status	85 points
CAAT	5	1.3% / 2.0%	2.0%	With Spouse: 60% J&S Without Spouse: LG5	75% (pre-2008 benefits)	85 points, or age 60 + 20 years
HOOPP	5	1.5% / 2.0%	2.0%	With Spouse: 60% J&S Without Spouse: LG15	75% (pre-2006 benefits)	age 60, or age 55 + 30 years
OMERS	5	1.325% / 2.0%	2.0%	With Spouse: 66-2/3% J&S Without Spouse: ROC	100%	age 55 + 30 years, or age 55 + 90 points
PSPP	5	1.3% / 2.0%	2.0%	With Spouse: 50% J&S Without Spouse: ROC	100%	90 points, or age 60 + 20 years

Source Table 6—Summary of Benefits for University DB Pension Plans and Public Sector Pension Plans
Final Report of the Working Group on University Pension Plans
Council of Ontario Universities, February 2010

Table 2. Comparison of Ontario DB Plan Contributions by Parties

DB Pension Plan	Average Earnings (yrs.)	Benefit Rate After Age 65	Benefit Rate Before Age 65	Subsidized Payment Form	Automatic Indexing (% of CPI)	Eligibility for Unreduced Early Retirement
University of Toronto						
Faculty	3	1.5% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	75%	age 60 + 10 years
Staff	3	1.6% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	75%	age 60 + 80 points
University of Ottawa ¹	5	1.3% / 2.0%	same	With Spouse: 60% J&S with G5 Without Spouse: LG5	CPI < 3%: 100% up to 2% CPI ≥ 3%: CPI – 1 %	age 60, or 90 points
McMaster University (Salaried Plan)	4	1.4% / 2.0%	same	With Spouse: 50% J&S Without Spouse: LG7	excess interest (threshold at 4.5%)	80 or 85 points by staff group
University of Waterloo	3	1.4% / 2.0%	same	LG10	100%	age 62
Ryerson University	5	1.35% / 2.0%	2.0%	With Spouse: 60% J&S Without Spouse: LG10	100%	90 points, or age 60 + 20 years
University of Guelph						
Professional Plan	3	1.5% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	CPI – 2%	age 55 + 85 points
Staff Plan	3	1.6% / 2.0%	same	With Spouse: 60% J&S Without Spouse: LG5	CPI – 2%	age 55 + 85 points
University of Windsor (Staff Plan)	5	1.5% / 2.0%	same	LG5	50% of excess interest (threshold at 6%) up to 50% of CPI	age 65

DB Pension Plan	Average Earnings (yrs.)	Benefit Rate After Age 65	Benefit Rate Before Age 65	Subsidized Payment Form	Automatic Indexing (% of CPI)	Eligibility for Unreduced Early Retirement
Trent University						
Faculty Plan	3	2.0%	same	With Spouse: 60% J&S Without Spouse: LG10	excess interest (threshold at 6.5%) up to 50% of CPI	age 65
Staff Plan	3	2.0%	same	With Spouse: 60% J&S Without Spouse: LG10	excess interest (threshold at 6.5%) up to 50% of CPI	age 60 + 80 points
OTPP	5	1.55% / 2.0%	2.0%	With Spouse: 50% J&S Without Spouse: LG10	100% (pre-2010 benefits); 50% (post-2009 benefits) plus top-up to 100% based on plan's funded status	85 points
CAAT	5	1.3% / 2.0%	2.0%	With Spouse: 60% J&S Without Spouse: LG5	75% (pre-2008 benefits)	85 points, or age 60 + 20 years
HOOPP	5	1.5% / 2.0%	2.0%	With Spouse: 60% J&S Without Spouse: LG15	75% (pre-2006 benefits)	age 60, or age 55 + 30 years
OMERS	5	1.325% / 2.0%	2.0%	With Spouse: 66-2/3% J&S Without Spouse: ROC	100%	age 55 + 30 years, or age 55 + 90 points
PSPP	5	1.3% / 2.0%	2.0%	With Spouse: 50% J&S Without Spouse: ROC	100%	90 points, or age 60 + 20 years

Source Table 6—Summary of Benefits for University DB Pension Plans and Public Sector Pension Plans
Final Report of the Working Group on University Pension Plans
Council of Ontario Universities, February 2010
<http://www.mcmaster.ca/mufa/COUPension%20Feb2010.pdf>

References

“Final Report of the Working Group on University Pension Plans”, Council of Ontario Universities, February 2010, <http://www.mcmaster.ca/mufa/COUPension%20Feb2010.pdf>.

“Rebuilding New Brunswick: The Case for Pension Reform”, A Government of New Brunswick Report, February 2013.

Simpson, Jeffrey, “New Brunswick Tackles its Pension Pickle”, The Globe and Mail (Toronto, ON), 02 June 2012: F.9.

Wright, Hugh, “New Brunswick Introduces Shared-Risk Pension Plans”, Canadian HR Reporter, 25. 17 (Oct 8, 2012): 18.

2012-2013 CAUT Almanac of Post-Secondary Education in Canada.

ASB 2013
Antigonish, Nova Scotia

Bobbi Morrison
St. Francis Xavier University
University of Strathclyde, PhD Candidate

Randy Delorey
St. Francis Xavier University
Saint Mary's University, PhD Candidate

I DIDN'T KNOW THAT: A CROSS-CULTURAL COMPARISON OF FACEBOOK USERS' PRIVACY LITERACY

The ubiquity of social network sites and the extensive information shared on these sites raises many questions about privacy. Despite the recognized importance of knowledge in consumer decision making, privacy knowledge has not been widely investigated and cross-cultural comparisons of this kind have not been undertaken. As such, the purpose of this paper is to explore and compare the privacy literacy of Facebook users in Canada and France in terms of subjective and objective privacy knowledge. Contrary to expectations, no significant differences were revealed between privacy literacy across cultures, yet descriptive results suggest many areas where privacy knowledge could be improved.

Introduction

Use of social network sites (SNS) has become ubiquitous. Facebook is the most popular social network site in the world with over one billion users worldwide (Fowler, 2012). Every day, users share more than 300 million photographs and click the 'like' button 2.7 billion times (Constine, 2012). This type of computer mediated communication is changing the way individuals communicate, subsequently raising questions about how businesses can harness the power of social interaction with their customer base (Archer-Brown, Piercy and Joinson, 2013) and provoking debate over privacy. Indeed, "privacy has become one of the most contested social issues of the information age" (Vasalou et al, 2012, p. 3, *from* Strickland and Hunt, 2005).

Generally, social network sites (SNS) are "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (Boyd and Ellison, 2007, 211). Personal information disclosure is a requisite part of SNS participation. Depending upon the privacy controls offered by the SNS and one's ability and choice to use such controls, the information shared can be viewed by SNS members' connections, connections-of-connections, open to the public, collected and used by the SNS provider and/or its corporate clients and countless silent listeners (Krishnamurthy and Wills, 2008; Stutzman et al, 2012; Bernstein et al, 2013).

Social network sites would not succeed without information disclosure, but it is this very characteristic that makes participants susceptible to privacy risks. Various claims made over the last decade cast speculation on whether individuals know enough about privacy to protect themselves in intensive information sharing environments (Cavoukian and Hamilton, 2002; Solove, 2006; Nissenbaum, 2010), yet few studies have attempted to measure privacy knowledge in order to assess whether these claims have merit. Furthermore, as privacy is recognized to be culturally relative (Westin, 1967; Moore, 2003), a determination of privacy knowledge in different cultural contexts is warranted to substantiate claims from a cross-cultural perspective, yet no such comparisons were identified in the literature. Thus, the overall objectives of this research were to investigate how much individuals knew about Facebook's privacy policies and to identify whether privacy literacy differed across cultures.

Literature Review

Privacy Literacy

There has been considerable research of late attempting to understand privacy concerns in online environments (i.e. Joinson et al, 2010; Xu et al, 2011; Midha, 2012), however, little research has been undertaken to investigate the role knowledge plays in understanding privacy in these environments. Since Nowak and Phelps asserted that attitudinal observations such as privacy concerns are 'only reliable and valid to the extent that consumers are knowledgeable and well-informed' about which they are concerned (Nowak and Phelps, 1992, 30), we maintain that understanding what one knows about privacy is essential to fully comprehend the nature of privacy in online environments. Furthermore, as knowledge has been shown to provide decision making control (Ajzen and Driver, 1991; Armitage and Conner, 1999; Awad and Krishnan, 2006; Chartrand, 2005) and affects individuals' behaviour (Buchanan et al, 2007) it follows that privacy knowledge should ultimately be a predictor of information disclosure decisions in SNS.

In the context of privacy, this knowledge has been referred to as 'privacy literacy', defined by Langenderfer and Miyazaki (2009) as "the understanding that consumers have of the information landscape with which they interact and their responsibilities within that landscape" (p. 383). As the self-disclosure of private information on SNS creates numerous instances where 'privacy leakage' may occur (Krishnamurthy and Wills, 2008), privacy literacy becomes an increasingly critical asset for consumers in an information landscape where the burden of privacy protection rests heavily in their hands (Langenderfer and Miyazaki, 2009; Nehf, 2007).

Consumer knowledge literature discusses two types of knowledge - subjective knowledge and objective knowledge (Alhabeeb, 2007; Brucks, 1985; Park and Lessig, 1981; Spreng, Divine and Page, 1990; Sujana, 1985). Subjective knowledge (SK) refers to the confidence one has in their knowledge about a subject, or what they think they know about something; whereas objective knowledge (OK) refers to the objective measurement of the accuracy of one's knowledge. Clearly, knowledge confidence (SK) and accuracy (OK) are related concepts (Ellen, 1994; Carlson et al 2009, 2007; Moorman et al 2004; Park Mothersbaugh and Feick, 1994), but they have

also been shown to be distinct constructs that impact consumer behaviour differently (Brucks 1985; Carlson et al 2009; Park et al, 1994).

Subjective knowledge. Notwithstanding the attention placed upon understanding the roles of SK and OK in consumer behaviour (i.e. Carlson et al, 2009), only one explicit (Morrison, 2013) and two implicit (Xu et al 2008; Lawler, 2012) measurements of SK were found within the privacy literature. Although Xu et al (2008) did not report conclusions about SK, Lawler, Molluzzo and Doshi (2012) and Morrison (2013) reported descriptive statistics indicating that confidence in privacy knowledge was not high. 'Awareness' has been mentioned in privacy investigations, but rather than being conceptualized as either SK or OK, this awareness referred to the subjective value consumers placed upon corporate information transparency (Malhotra et al, 2004; Awad and Krishnan, 2006; Krasnova and Veltri; 2010) or subjective self-assessments of social awareness (Dinev and Hart, 2006). In other instances, self-reported Internet experience, Internet literacy (Bellman et al, 2004; Dinev and Hart, 2006; Moscardelli and Divine, 2007) and perceived 'privacy self-efficacy' (Rifon, LaRose and Lewis, 2007; Tow, Dell and Venable, 2010) have been linked to privacy related decisions online. While these seemingly similar conceptualizations may be proxies for subjective knowledge, the association has not been conclusively determined. As such, there is little information about SNS user subjective privacy knowledge available for comparison.

Objective knowledge. Nissenbaum (2010) has stated that individuals are "not fully aware that at certain critical junctures information is being gathered or recorded. Nor do they fully grasp the implications of the information ecology in which they choose to act." (p.105). Indeed, several studies have articulated a sense that people remain unaware that their behaviour on SNS could be putting them at risk (Ofcom, 2008; Krishnamurthy and Wills, 2008). Given the strong theoretical connection between knowledge and behaviour, it is surprising that there has not been more attention to an assessment of objective privacy knowledge.

Studies within direct marketing have indicated that objective knowledge about marketing practices with privacy implications was low in most cases (Culnan, 1995; Foxman and Kilcoyne, 1993; Graeff and Harmon, 2002; Nowak and Phelps, 1992; Milne and Rohm, 2000) excepting one (Dommeyer and Gross, 2003). In online contexts, OK about website privacy policies (Acquisti and Grossklags, 2005; Hoofnagle and King, 2008; Turow et al 2005) and technical and legal privacy protections (Acquisti and Grossklags, 2005) was similarly revealed to be low for U.S. respondents. In fact, Hoofnagle et al (2010) concluded that the majority of Americans were 'privacy illiterate'. In Canada, Morrison (2013) revealed that objective knowledge about Canada's privacy legislation (*Personal Information and Protection of Electronic Documents Act, PIPEDA*) was collectively low and Harris Decima reported that just three in ten Canadians were aware of a federal institution that helps protect the privacy of their personal information (Harris Decima, 2011).

Reasons for low objective privacy knowledge have been attributed to consumer neglect in reading privacy policies (BusinessWire, 2010; Lawler, Molluzzo and Doshi, 2012; Winkler, 2001) or a failure to read them thoroughly (Office of National Statistics, 2011). This is perhaps due to the length and complexity of many online privacy policies which likely requires patience

most consumers do not have (Krashinsky and El Akkad, 2010). As a result, many people agree to privacy policy terms without genuinely knowing to what they have consented. To that end, Facebook users have regularly reported poor understanding of that site's privacy policies (Govani and Pashley, 2005; Gross and Acquisti, 2005; Acquisti and Gross, 2006) and erroneous assumptions (Ofcom, 2008) and clear misinterpretations (Turow, Hennessey and Bleakley, 2008) about the scope of privacy policies have been observed. Second, most SNS users neglect to make use of available privacy settings (Govani and Pashley, 2005; Gjoka et al 2011). And, even in instances where respondents claimed to understand the privacy policies of Facebook and make use of the privacy settings, researchers speculated that the weak criteria many use to accept 'friends' into their networks (which in turn allows information to be shared beyond one's assumed control) further signified poor privacy knowledge accuracy (OK) (Debatin et al, 2009).

Calibration. Calibration between what one thinks they know and what one accurately knows is an important consideration in consumer knowledge and critical to decision making because it allows consumers to cope with situations of incomplete information and errors (Alba and Hutchinson, 2000). It is meant to represent the agreement between objective and subjective knowledge assessments of the validity of information, but does not refer to the accuracy of information itself. When there is lack of agreement between knowledge measures, a miscalibration is noted. If confidence in knowledge is higher than knowledge accuracy, a consumer is referred to as overconfident; underconfidence refers to the opposite condition. Empirical contexts are rarely expected to achieve a high level of calibration, and some degree of over-confidence or under-confidence is more likely to occur. The privacy literature has speculated an optimistic miscalibration of privacy knowledge might exist among SNS users (Boyd, 2007; Debatin et al 2009; Livingstone 2008; Youn 2009). Morrison (2013) empirically detected a miscalibration of privacy knowledge among Canadian SNS users, but the direction of the miscalibration was neither consistently optimistic nor pessimistic.

Cultural Differences

Hofstede (1980) defined culture as "the collective programming of the mind which distinguishes members of one group or category from people from another" (p.260). Hofstede's Cultural Value Indices demonstrate cultural differences globally based upon five dimensions – power distance, individualism, uncertainty avoidance, masculinity and long-term orientation (1980; 2001) and this particular framework has become a popular way of examining cross-cultural differences (Kirkman, Lowe and Gibson, 2006). According to the most recent data compiled by the Hofstede Centre (Table 1), Canada and France exhibited key differences on certain dimensions of the cultural value index. First, uncertainty avoidance (UAI) in France was particularly strong (UAI = 86) and represented the strongest dimension in the scale meaning that the French culture is intolerant of unorthodox beliefs and behaviours. In contrast, Canadians (UAI = 48) tend to be more relaxed, less rule-oriented and less emotionally expressive. The next greatest discrepancy in cultural value index measures pertained to the dimension of Power Distance (PDI). Whereas French culture (PDI = 68) is accepting of hierarchical social structures wherein power is unequally distributed and information flows are determined by power, Canadians (PDI = 39) value egalitarianism and have a greater expectation of transparency in information flows. Although Masculinity scores for both countries were quite close (France MAS = 43; Canada MAS = 52),

France was considered a moderately feminine culture valuing quality of life and support of the underdog rather than competition in the workplace whereas Canada was considered a moderately masculine society with strong standards of performance but reserved attitudes toward success and winning. The Long-Term Orientation (LTO) dimension also exhibited differences between these two countries, but neither country had strong long-term outlooks. Long-Term Orientation scores for France (LTO = 39) and Canada (LTO = 23) indicated both cultures held short-term orientations respectful of tradition, valued normative behaviours and focused upon the pursuit of absolute truth and immediate gratification. Canada (IDV = 80) and France (IDV = 71) were shown to be culturally similar on the individualism dimension, indicating that people in each country hold strong expectations that the responsibility to care for oneself and immediate family lies with the individual rather than society at large. Indeed, in Canada, Individualism was noted to be the most prominent cultural value.

Table 1. Hofstede Cultural Indices for Canada and France¹

Country	Power Distance	Individualism	Uncertainty Avoidance	Masculinity	Long-term Orientation
Canada	39	80	48	52	23
France	68	71	86	43	39
World Avg ²	55	43	64	50	n/a

¹ Hofstede Centre ² In Veltri and Elgarah, 2009

According to Vishwanath (2004) there are two schools of thought about how culture and online behaviour may be conceptualized – mediated effects and culturally mediated interaction. With mediated effects, the argument is that the medium alters cultures such that online participation leads to global electronic villages characterized by cooperation (i.e. Hongladarom, 1998; Shah et al, 2001; Rheingold, 2002) or real-world societal erosion as evidenced by declining civic engagement (i.e. Shah et al, 2001). Also consistent with this perspective is the notion that alteration could occur at the level of the medium. As Qiu, Lin and Leung (2012) revealed, SNS users could switch their accustomed online behaviours to match those deemed appropriate in alternative online communities, thereby suggesting that the medium and participant group norms were important determinants of SNS behaviour and DeAndrea, Shaw and Levine (2010) surmised that Facebook might socialize self-expression.

On the other hand, the culturally mediated perspective argues that the normative effects of culture constrain how an individual uses various media (i.e. Grace-Farfaglia et al, 2006). This latter viewpoint has guided this research as there have been numerous examples wherein cross-cultural differences in both online and privacy behaviours have been noted. Specifically, Grace-Farfaglia et al (2006) concluded that although there was a technological convergence among societies, a global cultural village was not observed. Similarly, cross-cultural differences have been observed regarding the type of information disclosed on Facebook (Karl, Peluchette and Schlaegel, 2010; Auter and Elmrasy, 2012), privacy concerns in Facebook (Tsoi and Chen, 2011) and instances of deceptive behaviour in SNS (Lewis and George, 2008). Interestingly, Marshall et al (2008) also revealed cross-cultural differences in the types of information disclosed, but it was in the opposite direction hypothesized. The culturally relative nature of privacy has been well articulated (Westin, 1967; Moore, 2003). And, in online environments, cross-cultural differences at

the national level have been detected in the ways people perceive and respond to online privacy (Milberg et al, 1995; Dinev et al, 2006; Marshall et al, 2008; Cho, Rivera-Sánchez and Lim, 2009). Yet, some authors claim that cross-cultural privacy issues related to SNS have not received adequate attention in the academic literature (Ur and Wang, 2013). Indeed, the literature review conducted herein did not reveal any cross-cultural comparison of privacy literacy, suggesting that this topic is among those warranting exploration.

Research Hypotheses

The premise of this investigation is that privacy literacy about social network sites will be distinguished by national culture. While no complete comparison of cross-cultural privacy literacy was found, Harris, Van Hove and Lievens (2003) did reveal differences in subjective Internet knowledge between Belgian and American respondents in a study pre-dating social network site dominance. The paucity of research in this area provides little evidence upon which to substantiate a hypothesis about whether France or Canada will have higher subjective knowledge, it is nonetheless hypothesized that:

H1: There will be a significant difference in subjective knowledge between Canadian and French survey respondents.

With respect to objective knowledge, it is expected that respondents from France will be likely to delegate the authority of protecting privacy to government given the protectionist stance traditionally adopted by government in that country (Warlaumont, 2010) and supported by the high power distance index for France (Hofstede Centre). With these cultural orientations, there would be little need for French SNS users to necessarily possess a great deal of objective privacy knowledge as the belief would be that there would be adequate state controlled policies in place for individuals' protection. Conversely, Canada's comparatively low power distance index in conjunction with the higher individualism index (Hofstede Centre), suggests Canadians should possess more objective knowledge about privacy. Although Morrison (2013) revealed that objective knowledge about Canadian privacy legislation was low, Hofstede's Canadian PDI and IDV cultural values suggest that the relative importance of information transparency and the need for Canadians to protect themselves rather than rely on government protection, might have had something to do with that finding. As a result, it is expected that Canadians will be more proactive in independently seeking information about privacy on the SNS with which they interact and should therefore have higher objective knowledge than people in France. Although, evidence suggests that objective privacy knowledge among Canadians may still not be high. Thus, it is hypothesized that:

H2: Canadians will have higher objective knowledge about privacy than their French counterparts.

Since calibration reflects one kind of comparison between subjective and objective knowledge and cultural differences have been hypothesized for each of those measures, it follows that:

H3: There will be a significant difference in calibration between Canadian and French survey respondents.

Methodology

An embedded mixed methods research design, appropriate for investigations requiring instrument development (Creswell and Plano Clark, 2011), was employed in this research. Accordingly, focus groups served as a supplemental qualitative strand embedded within a dominant quantitative cross-sectional survey design. The focus group technique was utilized for the sole objective of survey development. Objective knowledge has typically been measured via a series of true/false questions (i.e. Carlson, Bearden and Hardesty, 2007) about a subject of interest but as a result, scales of this kind do not translate among contexts. As the only similar objective knowledge privacy measure (Morrison, 2012) identified by the researchers was created for the purposes of measuring consumer knowledge about Canadian privacy legislation, a new scale was required to assess objective knowledge about Facebook's privacy policy.

To develop this survey instrument, researchers crafted 29 statements from Facebook's data use policy (aka privacy policy) that were either true or false. A small focus group of undergraduate students discussed each statement in detail to assess statement clarity, comprehensibility and answerability. Based on focus group feedback, the researchers independently selected items that should remain on the survey instrument. Upon comparison, researchers had selected 18 items in common and were thus retained. Discussion among the researchers resulted in an additional 4 items being removed from the survey to avoid repetition. A second focus group was then held with another group of students to re-assess the clarity of the remaining items and to pre-test the survey. Once the survey instrument was finalized in English, professional translation services were obtained to complete an English to French translation of all text.

A quota sampling technique was employed for data collection to capture responses from Facebook users in Canada and France according to published age distribution characteristics of the Facebook population in each country. The services of Survey Sampling International (SSI), a professional data collection firm, were contracted to administer the survey online and ensure 200 completed responses from Canada (in English) and 200 completed responses (in French) from France were collected from individuals ranging in age from 18-64. SSI managed all correspondence with the participants. SSI provides incentives to its panel participants based on a points system that may be redeemed for rewards. The survey was hosted through Fluidsurveys.com and was conducted in July, 2013. Once data collection was complete, data was cleaned and responses from each country were combined into a single SPSS data file for analysis using SPSS v. 20.

Measures

Subjective knowledge (SK). The SK scale consisted of three items modified from Carlson, Bearden and Hardesty (2007) to reflect the context of privacy. On a 7-point Likert scale, respondents rated their perceived subjective privacy knowledge to three questions: 'Compared

with most people you know, how would you rate your knowledge about how Facebook collects and manages your personal information?' (1=One of the least knowledgeable to 7 = One of the most knowledgeable); 'In general, I am quite knowledgeable about how Facebook collects and manages my personal information.' (1 = Strongly Disagree to 7 = Strongly Agree); and 'I am quite knowledgeable about how all the information I provide in my online social network is collected and managed by Facebook.' (1 = Strongly Disagree to 7 = Strongly Agree). An SK score was computed as the average of the three subjective knowledge variables (min = 1, max = 7). To prevent bias, SK items were administered before OK items. The scale was determined to be internally consistent as Chronbach's α ($\alpha = 0.852$) exceeded the acceptable minimum of 0.7 (Nunnally, 1978).

Objective knowledge (OK). The OK measure was comprised of 14 True/False statements pertaining to Facebook's privacy policy. Respondents were given the option to select one of three possible answers to each question - true, false, or don't know. The scale was determined to be internally consistent as Chronbach's α ($\alpha = 0.839$) exceeded the recommended threshold of 0.70 (Nunnally, 1978). A total OK score was computed by scoring correct answers as '1' and incorrect or don't know answers as '0' and summing the results. Thus, OK scores could potentially range from 0 (none correct) to 14 (all correct).

Calibration. Calibration scores (CAL) were calculated by scoring OK responses to distinguish inaccurate answers. Since calibration is meant to reflect the difference between what one thinks they know and what one accurately knows, CAL was computed from the types of responses provided to the OK scale where a correct answer signified perfect calibration with knowledge accuracy and confidence on that item and an inaccurate answer signified overconfidence. Thus, to compute CAL, correct responses were assigned a value of 1, inaccurate answers were assigned a value of -1, and admissions of ignorance ('don't know') were assigned a value of 0. Thus, the theoretical range for the calibration score was -14 (all wrong) to +14 (all correct).

Results

The survey returned 288 responses from France and 265 from Canada. Upon review and data cleaning, 199 responses from France and 198 responses from Canada were retained for analysis. This represented a 69% and 75% completion rate for France and Canada, respectively.

The demographic distribution, by sex, of the sample was 151 (38%) male, 237 (60%) female and 9 (2%) preferred not to answer. Specifically, the gender distribution of the Canadian sample was 69 (35%) male, 124 (63%) female, and 5 (2%) preferred not to answer. The French sample was 82 (41%) male, 113 (57%) female, and 4 (2%) preferred not to answer. The gender distributions of survey respondents from each country were slightly more female than published Facebook membership distributions suggested where a 50/50 male/female split was noted in France and a 46% male to 54% female ratio was identified in Canada (Socialbakers, 2013a; 2013b). Table 2 displays the target distribution of the sample by age range, based upon Facebook user age profiles (Socialbakers, 2013a; 2013b) along with the actual sample distribution

achieved. As shown, the youngest age range (18-24) was slightly under-represented whereas each of the older age groups were slightly over-represented according to reported distributions.

Table 2. Target and Actual Sample Distributions by Age and Country

Age	Target Sample Distribution based on FB User Distributions ³				Actual Sample Distribution			
	FB Canada		FB France		Canada		France	
	N	%	N	%	#	%	#	%
18 - 24	52	25.9	59	29.4	33	16.7	33	16.6
25 - 34	56	28.1	61	30.6	57	28.8	66	33.2
35 - 44	40	20.2	40	20.0	49	24.7	52	26.1
45 - 54	31	15.7	26	12.9	35	17.7	31	15.6
55 - 64	20	10.1	14	7.1	24	12.1	17	8.5
	200	100.0	200	100.0	198	100.0	199	100.0

³ Socialbakers, 2013a; 2013b.

Table 3 provides a summary of results to the privacy objective knowledge (OK) assessment, pertaining to Facebook's privacy policy. What is immediately telling in these results is that only two questions achieved 50% correct and many questions resulted in an equal or greater percentage of respondents indicating they do not know the answer than who got the correct answer. Taken together these results suggest Facebook users' knowledge of Facebook's privacy policy is very low.

Table 3: Privacy Objective Knowledge Summary by Question

Variable	Description (Correct Answer)	Combined		Canada		France	
		% Correct	% Don't Know	% Correct	% Don't Know	% Correct	% Don't Know
OK 1	The only data collected by Facebook about you is what you provide directly to Facebook, such as your profile details, status updates, and comments. (FALSE)	27.8	16.9	30.5	21.2	25.1	12.6
OK 2	When you delete your account Facebook deletes all information that they've collected about you. (FALSE)	52.1	29.0	58.1	28.3	46.2	29.6
OK 3	Facebook owns all information it collects about you. (FALSE)	20.4	24.9	17.2	23.7	23.6	26.1
OK 4	Facebook provides guarantees about the accuracy and security of my personal information in their Data Use Policy (aka Privacy Policy). (FALSE)	22.7	31.7	25.3	33.3	20.1	30.2
OK 5	I can control whether or not public information on my Facebook Timeline may be made visible to third party search engines (ie. Google or Bing). (TRUE)	50.3	29.0	54	23.2	46.7	34.7
OK 6	Facebook controls, and is responsible for, all data collected about me through Facebook Apps that I use. (FALSE)	29.7	28.0	32.8	26.8	26.6	29.1
OK 7	Sometimes, Facebook Facebook advertisers post promoted news stories to timelines with your personal information. When this happens, the advertisers are violating Facebook policy for use of your personal information. (FALSE)	22.4	43.3	22.2	43.9	22.6	42.7
OK 8	Only information you have marked as "Public" will be accessible to people other than your "Friends". (FALSE)	17.9	12.6	16.7	12.1	19.1	13.1
OK 9	When you set a post to display to "Friends" only your friends will ever be able to see the contents of that post. (FALSE)	19.4	11.4	20.2	11.1	18.6	11.6
OK 10	I can limit access to my Facebook profile name, profile picture, and cover photo in my Facebook Profile Privacy settings. (FALSE)	16.9	14.6	17.7	12.6	16.1	16.6
OK 11	Facebook Pages are always public, and so anything posted to the Page is also always public. (TRUE)	25.7	18.9	23.2	19.7	28.1	18.1
OK 12	If I want to prevent Friends from tagging me in a post or photo I can update my privacy settings in Facebook. (FALSE)	14.6	29.5	13.1	26.3	16.1	32.7
OK 13	Once tagged by a Friend in a photo or post I cannot remove the tag. (FALSE)	46.6	24.7	53	22.7	40.2	26.6
OK 14	Regardless of your privacy settings, anyone who uses Facebook may tag you in a photo or post, even if they are not a Facebook Friend. (FALSE)	31.2	34.3	33.8	36.4	28.6	32.2

Table 4 contains the descriptive results of the privacy knowledge scores for subjective privacy knowledge (SK), objective privacy knowledge (OK), and a calibrated objective privacy knowledge (CAL) score. Recall that OK and CAL are based upon fourteen objective questions pertaining to Facebook’s privacy policy and SK was measured on a 7 point Likert scale. SK was essentially neutral in Canada ($SK_{\text{Canada}} M = 3.98, SD = 1.42$) and slightly lower in France ($SK_{\text{France}} M = 3.77, SD = 1.42$). Results partially support the second hypothesis that OK would be higher in Canada, although OK was notably low in both cultural contexts ($OK_{\text{France}} M = 3.78, SD = 2.56$; $OK_{\text{Canada}} M = 4.18, SD = 2.41$). Descriptive statistics for CAL revealed an overall miscalibration of privacy knowledge ($CAL_{\text{France}} M = -2.83, SD = 3.90$; $CAL_{\text{Canada}} M = -2.21, SD = 3.67$).

Table 4: Privacy Knowledge Descriptive Results

	Combined			Canada			France		
	SK	OK	CAL	SK	OK	CAL	SK	OK	CAL
N	393	397	397	197	198	198	196	199	199
Mean	3.87	3.98	-2.52	3.98	4.18	-2.21	3.77	3.78	-2.83
Std. Deviation	1.42	2.49	3.80	1.42	2.41	3.67	1.42	2.56	3.90
Minimum	1	0	-10	1	0	-10	1	0	-10
Maximum	7	12	10	7	12	10	7	11	9
Theoretical Min	1	0	-14	1	0	-14	1	0	-14
Theoretical Max	7	14	14	7	14	14	7	14	14

To assess for differences in privacy literacy between Canada and France, an ANOVA, using SPSS v.20, was conducted for each of the three privacy knowledge measures. Results are presented in Table 5. As can be seen, the ANOVA assessment of the subjective knowledge score between countries resulted in a non-significant result ($F(1, 391) = 2.13, p > 0.05$, partial $\eta^2 = 0.005$), thus H1 was not supported. Similarly, assessment of the objective knowledge score between countries resulted in a non-significant difference between cultures ($F(1, 395) = 2.54, p > 0.05$, partial $\eta^2 = 0.006$), thus H2 was rejected. H3 was also rejected as the ANOVA assessment of the objective knowledge calibrated score between countries resulted in a non-significant result ($F(1, 395) = 2.68, p > 0.05$, partial $\eta^2 = 0.007$).

Table 5: Univariate Analyses of Variance for Privacy Literacy

	SK ^a	OK ^a	CAL ^a
Country (F ratio)	2.13	2.54	2.68

a. Univariate $df = 1,391$ (country SK) and $1,395$ (country OK and country CAL)

Discussion

Despite conclusions within the literature that cross-cultural differences influenced privacy concerns and privacy behaviours (Milberg et al, 1995; Dinev et al, 2006; Marshall et al, 2008; Cho, Rivera-Sánchez and Lim, 2009), the same connection could not be made with privacy literacy. The results of this study clearly indicated there was not a significant difference between Facebook users' privacy literacy in Canada and France and as a result, none of the hypotheses could be supported. Specifically, there were no cross-cultural differences revealed in subjective knowledge, objective knowledge or calibration.

Rather than the culturally-mediated interaction perspective of cultural influence used to frame this study and formulate hypotheses, these results might be suggesting that mediated effects may be taking place within the technological environment of Facebook similar to the behaviour modification observed by Qiu, Lin and Leung (2012) and DeAndrea, Shaw and Levine (2010). However, as the results only compared privacy literacy across nations, evidence of cultural convergence consistent with expectations that a comprehensive global virtual community will emerge (i.e. Rheingold, 2002) has not been provided. Therefore, care must be taken to not generalize too broadly from these non-significant results.

It is possible that the lack of cross-national difference evidenced in this study could be due to the similarities among certain cultural values within the two countries. The comparison of Hofstede's cultural value indices for the two countries presented in Table 1 identified that both Canada and France were quite close on the individuality dimension and both were more individualistic nations than the world average ($IDV_{Canada} = 80$; $IDV_{France} = 71$; $IDV_{World} = 43$). This is an important consideration given Lewis and George's (2008) finding that it was *only* the IDV cultural value dimension that behaved as expected in their cross-cultural comparison of Koreans and Americans and Sondergaard's (1994) conclusion that more significant cross-cultural differences were associated with the IDV dimension. Thus, it is possible that the cultures of the selected nations were not distinct enough to return differences in privacy literacy. Therefore future cross-cultural comparisons should endeavor to utilize samples from countries with dissimilar individualism values.

Likewise, it is also possible that the lack of cross-national privacy literacy difference could be explained by the similarities in the privacy regulatory environments of the two countries under investigation. Although it was hypothesized that French respondents would have little need to possess great amounts of objective knowledge due to the nation's cultural power distance value relative to Canada and the World, ($PDI_{France} = 68$; $PDI_{Canada} = 39$; $PDI_{World} = 55$) and the likely expectation by SNS users that the country's protectionist policies (Warlaumont, 2010) would extend to the realms of social network sites and privacy, both countries have extensive privacy legislation that cover all information exchange interactions with businesses. In Canada, the Personal Information and Protection of Electronic Documents Act (PIPEDA) is overseen by the Office of the Privacy Commissioner of Canada; in France the European Union Data Protection Directive (Directive 95/46) is overseen by the Commission Nationale de l'Informatique et des Libertés (CNIL). So, on one hand, Canada's smaller power distance values might not be as important as the universal privacy legislation which could conceivably offer sufficient perceived privacy pro-

tection to SNS users and thus may minimize quests for objective knowledge. Of course, given previous findings that objective knowledge of Canadian privacy legislation was also low (Morrison, 2013), it is unlikely that Canadians' poor privacy literacy can be explained by confidence in the privacy regulatory environment acquired through objective knowledge about it. Thus, future research should attempt to clarify whether cross-cultural differences in privacy literacy might be detected between nations with different regulatory environments and should also consider other factors related to one's general confidence in the legal protections offered in their nations.

Despite the lack of cross-cultural differences revealed in the results, there is compelling evidence provided herein that privacy literacy is poor. From the descriptive results of the objective knowledge test (Table 3), we learned that in both Canada and France, respondents knew very little about the contents of Facebook's privacy policy, thereby reinforcing speculation that most people do not know enough to protect themselves in technological environments (Cavoukian and Hamilton, 2002; Solove, 2006; Krishnamurthy and Wills, 2008; Nissenbaum, 2010) with empirical evidence. From a consumer advocacy standpoint, this poor objective privacy knowledge is a discouraging finding especially in light of the overconfidence realized in the results from both countries. When asked how much they think they know, respondents from each nation reported close to a neutral subjective knowledge, suggesting that they believed they knew enough about privacy on Facebook to get by. However, calibrated scores of the objective knowledge test suggested otherwise. Clearly, respondents from each nation were more confident in their knowledge about Facebook's privacy policy than their objective knowledge would suggest ($CAL_{France} M = -2.83, SD = 3.90$; $CAL_{Canada} M = -2.21, SD = 3.67$).

There are several implications of these findings to various stakeholder groups. First, academic pursuits will be guided by the suggestions for future research identified in this discussion. In addition, future research should consider investigating what role poor privacy literacy has on information disclosure decisions on social network sites. Second, governments should be interested to learn how little people actually know about their privacy in the technological environments in which they regularly interact. As governments serve to protect its citizens, the collective evidence of poor privacy knowledge and the specific areas of privacy illiteracy identified herein may be used to inform education and training programs for the public. The results also have implications for businesses that operate social network sites and those that use them to engage customers. In one infamous example, Facebook was forced to eliminate 'Beacon' due to public outrage. The Beacon service publicly broadcast purchases that Facebook members had made online with external retailers in the Facebook Newsfeed. Although this practice did not contravene Facebook's privacy policy, Facebook members were angered by the perceived privacy violation and perceived deception by the company. As a result of the negative backlash, Facebook halted the Beacon service. Where it is conceivable that consumer perception might have been different if Facebook users had been knowledgeable about Facebook's privacy policy, it appears adequate privacy literacy might have prevented this problem and therefore SNS such as Facebook should consider ways to improve member's privacy knowledge. Similarly, third parties that use Facebook to connect with customers should be aware of the low levels of privacy literacy of members and seek ways to encourage improved privacy understanding so that they do not also engage in activities that are perceived to be inconsistent with privacy policies. Therefore, to prevent future similar mishaps, Facebook and other social network sites should consider tactics that would en-

courage users to both read and understand privacy policies including addressing issues such as accessibility, length and readability.

Limitations

Some caution must be exercised in interpreting the results of this study due to a few important limitations associated with sampling, measurement and assumptions. In terms of sampling issues, there was likely some self-selection bias in the sample as a result of reliance upon panel participants incentivized through a points program. In addition, there was an unknown bias as the sample was limited to Facebook users, but there are no indicators to compare whether Facebook privacy knowledge is representative of the general population and/or non-Facebook users. Attention should also be directed to the demographic distribution of the sample achieved. Despite the quota approach, there were some minor deviations from the targeted population estimates in that the sample was slightly underrepresented in the 18-24 year age range and was slightly more female than desired. Among the measurement limitations were the answers to the objective knowledge quiz - all but two of the fourteen items on the objective knowledge scale had the correct answer of False. In addition, the Facebook privacy policy changes frequently so that what was true when the survey was deployed may not have been true in the past nor is it guaranteed to be true in the future. It is also important to note that OK was measured exclusively for Facebook; however, privacy knowledge extends well beyond this context. Furthermore, though a calibration measure was computed to provide insight into the consistency of privacy knowledge confidence and privacy knowledge accuracy measures, the calibration measure was not without limitations. The calibration measure calculated herein was only capable of indicating situations of calibration and overconfident miscalibration, but the measure could not detect situations wherein there was underconfidence. Furthermore, Hofstede's Cultural Value Indices were not directly measured for survey respondents, thus cultural values were only assumed to be consistent with published index values. Finally, a discriminant analysis would aid in further interpretation of cultural differences but was not undertaken in this study. Thus, further analysis of this data will incorporate discriminant tests.

Conclusions

The objective of this research was to investigate whether cross-cultural differences in privacy literacy could be observed among Facebook users from Canada and France. While results suggested there were no significant differences between cultures on measures of objective privacy knowledge, subjective privacy knowledge or calibration, interesting findings were revealed with respect to privacy literacy. Specifically, privacy literacy was concluded to be poor in both Canada and France as evidenced by poor objective knowledge and individuals' overly optimistic assessment of their privacy knowledge. Thus, it was suggested that SNS participants do not likely have the requisite ability to protect themselves from privacy violations in these environments. The results also suggest implications for future academic inquiry and opportunities for both government and business to improve privacy literacy levels among the public.

References

Alba, J.W., & Hutchinson, J.W. "Knowledge Calibration: What Consumers Know and What They Think They Know", *Journal of Consumer Research*, 27(2), 2000, 123-156.

Acquisti, A. & Grossklags, J., "Privacy and Rationality in Consumer Decision Making", *IEEE Security & Privacy*, (January/February 2005), 24-30. <http://csis.pace.edu/~ctappert/dps/d861-09/team2-3.pdf>. [Accessed July 12, 2011].

Acquisti, A., & Gross, R., "Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook" In *Privacy Enhancing Technologies Workshop (PET)*, 2006.

Alhabeeb, M.J., "Consumer Knowledge and Consumption: A Human Capital Model", *Academy of Marketing Studies Journal*, 11(1), 2007, 69-81.

Archer-Brown, C., Piercy, N. and Joinson, A., "Examining the Information Value of Virtual Communities: Factual Versus Opinion-Based Message Content," *Journal of Marketing Management*, 29(3-4), (2013), 421-438.

Armitage, C. J., & Conner, M., "The Theory of Planned Behavior: Assessment of Predictive Validity and 'Perceived Control'", *British Journal of Social Psychology*, 38(1), (1999), 35-54.

Auter, P.J. and Elmrazy, M., "A Cross-Cultural Content Analysis of Student Facebook Use," Paper presented at the 6th Annual Forum of the Saudi Association for Media and Communication, King Saud University, (April 15-16, 2012).

Awad, N.F. & Krishnan, M.D., "The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to be Profiled Online for Personalization", *MIS Quarterly*, 30(1), (2006), 13-28.

Azjen, I. and Driver, B. L. "Prediction of Leisure Participation from Behavioral, Normative, and Control Beliefs: An Application of the Theory of Planned Behavior," *Leisure Sciences* 13(3), (1991), 185-204.

Bellman, S., Johnson, E.J., Kobrin, S.J. and Lohse, G.L., "International Differences in Information Privacy Concerns: A Global Survey of Consumers," *The Information Society*, 20, (2004), 313-324.

Bernstein, M.S., Bakshy, E., Burke, M. and Karrer, B., "Quantifying the Invisible Audience in Social Networks," CHI, Paris, France, (April 27-May 2, 2013).

Boyd, d., "Why Youth Heart Social Network Sites: The Role of Networked Publics in Teenage Social Life", *The Berkman Center for Internet and Society Research Publication Series*, Harvard University, No. 2007-16, (2007), 119-142.

- Boyd, d. m., & Ellison, N. B., "Social Network Sites: Definition, History, and Scholarship", *Journal of Computer-Mediated Communication*, 13(1), (2007), 210-230, <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>. [Accessed January 7, 2011]
- Brucks, M., "The Effects of Product Class Knowledge on Information Search Behavior", *Journal of Consumer Research*, 12(June), (1985), 1-16.
- Buchanan, T., Paine, C., Joinson, N. & Reips, U.D., "Development of Measures of Online Privacy Concern and Protection for Use on the Internet," *Journal of the American Society for Information Science and Technology*, 58(2), (2007), 157-165.
- Business Wire, "The Truth About Social Media Identity Theft: Perception Versus Reality", (June 21, 2010), <http://www.businesswire.com/news/home/20100621005370/en/Truth-Social-Media-Identity-Theft-Perception-Reality>. Accessed June 7, 2012.
- Carlson, J.P., Bearden, W.O., & Hardesty, D.M., "Influences on What Consumers Know and What They Think They Know Regarding Marketer Pricing Tactics", *Psychology & Marketing*, 24(2), (2007), 117-142.
- Carlson, J.P., Vincent, L.H., Hardesty, D.M. & Bearden, W.O., "Objective and Subjective Knowledge Relationships: A Quantitative Analysis of Consumer Research Findings", *Journal of Consumer Research*, 35 (February), (2009), 864 – 876.
- Cavoukian, A. & Hamilton, T.J., *The Privacy Payoff: How Successful Businesses Build Customer Trust*, Toronto ON: McGraw-Hill Ryerson, 2002.
- Chartrand, T.L., "The Role of Conscious Awareness in Consumer Behavior", *Journal of Consumer Psychology*, 15(3), (2005), 203-210.
- Cho, H., Rivera-Sánchez, M. and Lim, S.S., "A Multinational Study of Online Privacy: Global Concerns and Local Responses," *New Media and Society*, 11(3), (2003), 395-416.
- Constine, J. "How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day", *Techcrunch.com* [web page] (2012); <http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/>. Accessed July 29, 2013.
- Creswell, J. W. and Plano Clark, V.L., *Designing and Conducting Mixed Methods Research*, 2nd ed., Thousand Oaks, CA: Sage Publications, 2011.
- Culnan, M.J., "Consumer Awareness of Name Removal Procedures: Implications for Direct Marketing," *Journal of Direct Marketing*, 9 (2), (1995), 10-19,

DeAndrea, D.C., Shaw, A.S. and Levine, T.R., "Online Language: The Role of Culture in Self-Expression and Self-Construal on Facebook," *Journal of Language and Social Psychology*, 29(4), (2010), 425-442.

Debatin, B., Lovejoy, J.P., Horn A.K., & Hughes, B.N., "Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences", *Journal of Computer-Mediated Communication*, 15, (2009), 83-108.

Dinev, T. and Hart, P., "Internet Privacy Concerns and Social Awareness as Determinants of Intention to Transact," *International Journal of Electronic Commerce*, 10(2), (2006), 7-29.

Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I., and Colautti, C., "Privacy Calculus Model in E-Commerce – A Study of Italy and the United States," *European Journal of Information Systems*, 15, (2006), 389-402.

Dommeyer, C.J. & Gross, B.L., "What Consumers Know and What They Do: An Investigation of Consumer Knowledge, Awareness, and Use of Privacy Protection Strategies", *Journal of Interactive Marketing*, 17(2), (2003), 34-51.

Ellen, P.S., "Do We Know What We Need to Know? Objective and Subjective Knowledge Effects on Pro-Ecological Behaviors," *Journal of Business Research*, 30(1), (1994), 43-52.

Fowler, G.A., "Facebook: One Billion and Counting", *The Wall Street Journal* [online] October 4, 2012; <http://online.wsj.com/article/SB10000872396390443635404578036164027386112.html>. Accessed July 29, 2013.

Foxman, E. R., & Kilcoyne, P., "Information Technology, Marketing Practice, and Consumer Privacy: Ethical Issues," *Journal of Public Policy & Marketing*, 12(1), (1993), 106-119.

Gjoka, M., Kurant, M., Butts, C.T., & Markpoulou, A., "A Walk in Facebook: Uniform Sampling of Users in Online Social Networks", *IEEE Journal on Selected Areas in Communications (JSAC)*, Special Issue on Measurement of Internet Topologies, (2011).

Govani, T. & Pashley, H., "Student Awareness of the Privacy Implications When Using Facebook", (2005), <http://lorrie.cranor.org/courses/fa05/tubzhlp.pdf>. Accessed May 12, 2012].

Grace-Farfaglia, P., Dekkers, A., Sundararajan, B., Peters, L., Park, S.-H., "Multinational Web Uses and Gratifications: Measuring the Social Impact of Online Community Participation Across National Boundaries," *Electronic Commerce Research*, 6, (2006), 75-101.

Graeff, T.R. & Harmon, S., "Collecting and Using Personal Data: Consumers' Awareness and Concerns", *Journal of Consumer Marketing*, 19(4), (2002), 302 – 318.

Gross, R., & Acquisti, A. "Information Revelation and Privacy in Online Social Networks", Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, 71-80, (2005), DOI [10.1145/1102199.1102214](https://doi.org/10.1145/1102199.1102214)

Harris Decima, "2011 Canadians and Privacy Survey", Presented to the Office of the Privacy Commissioner of Canada, (March 31, 2011), http://www.priv.gc.ca/information/por-rop/2011/por_2011_01_e.pdf. Accessed August 25, 2011.

Harris, M.M., Van Hove, G. and Lievens, F., "Privacy and Attitudes Toward Internet-Based Selection Systems: A Cross-Cultural Comparison," International Journal of Selection and Assessment, 11(2/3), (2003), 230-236.

Hofstede Centre, <http://geert-hofstede.com>. Accessed June 29, 2013.

Hofstede, G., Culture's Consequences: International Differences in Work-Related Values. Beverly Hills CA: Sage, 1980.

Hofstede, G., Culture's consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Thousand Oaks CA: Sage Publications, 2001.

Hongladarom, S., "Global Culture, Local Culture, and the Internet: The Thai Example," Electronic Journal Communication, 8(3-4), (1998).

Hoofnagle, C. J. & King, J., "Research Report: What Californians Understand About Privacy Online", (September 3, 2008), http://www.truste.com/pdf/09032008_BerkeleyStudy.pdf. Accessed July 20, 2012.

Joinson, A.N., Paine, C., Reips, U.-D., and Buchanan, T., "Privacy, Trust and Self-Disclosure Online," Human-Computer Interaction, 25, (2010), 1-24.

Karl, K., Peluchette, J. and Schlaegel, C., "Who's Posting Facebook Faux Pas? A Cross-Cultural Examination of Personality Differences," International Journal of Selection and Assessment, 18(2), (2010), 174-186.

Kirkman, B., Lowe, K. and Gibson, C. "A Quarter Century of *Culture's Consequences*: A Review of Empirical Research Incorporating Hofstede's Cultural Values Framework," Journal of International Business Studies, 37(3), (2006), 285-320.

Krashinsky, S. & El Akkad, O., "The End of Online Privacy", Globe and Mail, (August 13, 2010), <http://www.theglobeandmail.com/news/technology/the-end-of-online-privacy/article1672466/page3/>. Accessed on May 11, 2011.

Krasnova, H. & Veltri, N.F., "Privacy Calculus on Social Networking Sites: Explorative Evidence from Germany and USA", Proceedings of the 43rd International Conference on System Sciences, Hawaii, (2010), 1-10.

Krishnamurthy, B. and Wills, C., "Characterizing Privacy in Online Social Networks," Proceedings of WOSN '08, (August 18, 2008), Seattle, Washington, 37-42.

Langenderfer, J. & Miyazaki, A. D., "Privacy in the Information Economy", *Journal of Consumer Affairs*, 43(3), (2009), 380-388.

Lawler, J.P., Molluzzo, J.C. & Doshi, V., "An Expanded Study of Net Generation Perceptions on Privacy and Security on Social Networking Sites (SNS)", *Information Systems Education Journal*, 10(1), (2012), 21-36.

Lewis, C.C. and George, J.F., "Cross-Cultural Deception in Social Networking Sites and Face-to-Face Communication," *Computers in Human Behavior*, 24, (2008), 2945-2964.

Livingstone, S., "Taking Risky Opportunities in Youthful Content Creation: Teenagers' Use of Social Networking Sites for Intimacy, Privacy and Self-Expression," *New Media & Society*, 10(3), (2008), 393-411.

Malhotra, N. K., Kim, S.S. & Agarwal, J., "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale and a Causal Model", *Information Systems Research*, 15(4), (2004), 336-355.

Marshall, B. A., Cardon, P. W., Norris, D. T., Goreva, N., & D'Souza, R., "Social Networking Websites in India and the United States: A Cross-National Comparison of Online Privacy and Communication," *Issues in Information Systems*, IX(2), (2008), 87-94.

Midha, V., "Impact of Consumer Empowerment on Online Trust: An Examination Across Genders," *Decision Support Systems*, 54(1), (2012), 198-205.

Milberg, S.J., Burke, S.J, Smith, H.J. and Kallman, E.A., "Values, Personal Information Privacy and Regulatory Approaches," *CACM*, 38(12), (December, 1995), 65-74.

Milne, G. R. & Rohm, A.J., "Consumer Privacy Name Removal Across Direct Marketing Channels: Exploring Opt-in and Opt-out Alternatives", *Journal of Public Policy and Marketing*, 19(2), (2000), 238-249.

Moore, A.D., "Privacy: Its Meaning and Value," *American Philosophical Quarterly*, 40(3), (2003), 215-227.

Moorman, C., Diehl, K., Brinberg, D., and Kidwell, B., "Subjective Knowledge, Search Locations, and Consumer Choice," *Journal of Consumer Research*, 31 (December), (2004), 673-80.

Morrison, B., "Do We Know What We Think We Know? An Exploration of Online Social Network Users' Privacy Literacy," *The Workplace Review*, (April, 2013), 58-79.

Moscardelli, D.M. and Divine, R., "Adolescents' Concern for Privacy When Using the Internet: an Empirical Analysis of Predictors and Relationships with Privacy Predicting Behaviors," *Family and Consumer Sciences Research Journal*, 35(3), (2007), 232-252.

Nehf, James P., "Shopping for Privacy on the Internet", *Journal of Consumer Affairs*, 41 (Winter), (2007), 351-365.

Nissenbaum, H. F., *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law and Politics, 2010.

Nowak, G.J. & Phelps, J., "Understanding Privacy Concerns: An Assessment of Consumers' Information-Related Knowledge and Beliefs", *Journal of Direct Marketing*, 6(4), (1992), 28-39.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Ofcom, "Social Networking: A Quantitative and Qualitative Research Report into Attitudes Behaviour and Use", Government Office of Communications, United Kingdom, (2008).

Office for National Statistics, "Statistical Bulletin: Internet Access – Households and Individuals, 2011", Office for National Statistics, (August 31, 2011).

Park, C. & Lessig, V.P., "Familiarity and its Impact on Consumer Decision Biases and Heuristics", *Journal of Consumer Research*, 8 (September), (1981), 223-230.

Park, C.W., Mothersbaugh, D.L., & Feick, L., "Consumer Knowledge Assessment", *Journal of Consumer Research*, 21(June), (1994), 71-82.

Qiu, L., Lin, H. and Leung, A.K.-Y., "Cultural Differences and Switching of In-Group Sharing Behavior Between an American (Facebook) and a Chinese (Renren) Social Networking Site," *Journal of Cross-Cultural Psychology*, XX(X), (2012), 1-16.

Rifon, N. J., LaRose, R. J., and Lewis, M. L., "Resolving the Privacy Paradox: Toward a Social-Cognitive Theory of Consumer Privacy Protection," Mimeo, Michigan State University, [online] <https://www.msu.edu/~wirthch1/privacyparadox07.pdf>. Accessed July 12, 2013.

Rheingold, H. *Smart Mobs: The Next Social Revolution*. Cambridge MA: Perseus Book Group, 2002.

Shah, D.V., Kwak, N. and Holbert, R.L., "Connecting and Disconnecting with Civic Life: Patterns of Internet Use and the Production of Social Capital," *Political Communication*, 18, (2001), 141-162.

Spreng, R. A., Divine, R.L., & Page, T., "An Empirical Examination of the Differential Effects of Objective and Subjective Knowledge on Information Processing", *American Marketing Association Conference Proceedings*, 12, (2001), 329-335.

- Socialbakers, "Canada Facebook Statistics." [web page], 2013a, <http://www.socialbakers.com/facebook-statistics/canada>. Accessed June 6, 2013.
- Socialbakers. "France Facebook Statistics." [web page], 2013b, <http://www.socialbakers.com/facebook-statistics/france>. Accessed June 6, 2013.
- Solove, D., "A Taxonomy of Privacy," *University of Pennsylvania Law Review*, 154(3), (2006), 477. [online] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=667622. Accessed August 1, 2011.
- Sondergaard, M., "Research Note: Hofstede's Consequences: A Study of Reviews, Citations and Replications," *Organization Studies*, 15(3), (1994), 447-456.
- Strickland, L.S. and Hunt, L.E., "Technology, Security, and Individual Privacy: New Tools, New Threats, and New Public Perceptions," *Journal of the American Society for Information Science and Technology*, 56(3), (2005), 221-234.
- Stutzman, F., Vitak, J., Ellison, N. B., Gray, R., and Lampe, C., "Privacy in Interaction: Exploring Disclosure and Social Capital in Facebook. *In International Conference on Weblogs and Social Media (ICWSM'12)*, Dublin, IE, (2012).
- Sujan, M., "Consumer Knowledge: Effects on Evaluation Strategies Mediating Consumer Judgments", *Journal of Consumer Research*, 12(June), (1985), 31-46.
- Tow, W. N. F. H., Dell, P., and Venable, J., "Understanding Information Disclosure Behaviour in Australian Facebook Users," *Journal of Information Technology*, 25(2), (2010), 126-136.
- Tsoi, H.K. and Chen, L., "From Privacy Concern to Uses of Social Network Sites: A Cultural Comparison via User Survey," *Proceedings of IEEE International Conference on Privacy, Security, Risk and Trust*, (2011), 457-464.
- Turow, J. Feldman, L. & Meltzer, K., "Open to Exploitation: American Shoppers Online and Offline", *Annenberg Public Policy Center of the University of Pennsylvania*, (Jun, 1, 2005), <http://www.annenbergpublicpolicycenter.org/NewsDetails.aspx?myId=31>. Accessed May 23, 2012.
- Turow, J., Hennessy, M. & Bleakley, A., "Consumers' Understanding of Privacy Rules in the Marketplace", *Journal of Consumer Affairs*, 42(3), (2008), 411-424.
- Ur, B. and Wang, Y., "A Cross-Cultural Framework for Protecting User Privacy in Online Social Media," *WWW 2013 Companion*, (May 13-17, 2013).

Vasalou, A., Gill, A. J., Mazanderani, F., Papoutsis, C. and Joinson, A., "Privacy dictionary: a new resource for the automated content analysis of privacy," *Journal of the American Society for Information Science and Technology*, 62 (11), (2011), 2095-2105.

Veltri, N.F. and Elgarah, W., "The Role of National Cultural Differences in User Adoption of Social Networking," *Proceedings of the Southern Association for Information Systems Conference*, Charleston, SC, (March 12-14, 2009), 169-174.

Vishwanath, A., "Manifestations of Interpersonal Trust in Online Interaction: A Cross-Cultural Study Comparing the Differential Utilization of Seller Ratings on eBay Participants in Canada, France and Germany," *New Media and Society*, 6(2), (2004), 219-234.

Warlaumont, H.G., "Review Article: Social Networks and Globalization: Facebook, YouTube, and the Impact of Online Communities on France's Protectionist Policies," *French Politics*, 8(2), (2010), 204-214.

Westin, A. F., *Privacy and Freedom*. New York: Atheneum, 1967.

Winkler, S., "Privacy: A Modern Challenge", City of Cambridge, MA Consumer's Council, (2001), <http://www.ci.cambridge.ma.us/consumer/Privacy1.html>. [Accessed June 14, 2011].

Xu, H., Dinev, T., Smith, H.J., & Hart, P., "Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View", *Proceedings of the 29th International Conference on Information Systems (ICIS)*, (2008), 1-16.

Xu, H., Dinev, T. Smith, H.J. Hart, P., "Information Privacy Concerns: Linking Individual Perception with Institutional Privacy Assurances," *Journal of the Association of Information Systems*, 12(12), (2011), 798-824.

Youn, S., "Determinants of Online Privacy Concern and its Influence on Privacy Protection Behaviors Among Young Adolescents," *The Journal of Consumer Affairs*, 43(3), (2009), 389-418.

WHY DO COMMITTED CUSTOMERS DEFECT? THEY ARE NOT ENGAGED!

Abstract

Marketing practitioners and academicians have been talking a lot about customer engagement in recent years (Sprott et.al, 2009; Fournier, 2011). By all accounts it is the newest and hottest “construct” in the field these days (Bolton, 2011). But, what is it? Is it a construct in and of itself? Is it a model or set of relationships? Is it some type of second (or third) order factor that subsumes a number of other constructs in the field? Or is it something else? This paper will focus on the “something else” explanation for customer engagement. Specifically, it will put forward that engagement might be little more than a specific type relationship between commitment and loyalty.

The Nature of Customer Engagement

The study of engagement in organizational settings is complicated because it appears to be a mish-mash of previously well studied concepts and relationships (Macey and Schneider, 2008). In the organizational behavior literature, when employee engagement has been measured as a distinct construct, it has been highly correlated with other key constructs such as job satisfaction and commitment to the organization (Saks, 2006). However, a common view is that it is significantly rooted in employee attitudes about various aspects of the organization for which they work and that these attitudes drive important behaviours in the workplace.

In the discipline of marketing, there is also a dominant view that customer engagement is ripe with both cognitive and attitudinal components (Sprott et.al, 2009; Brodie et.al, 2011). Evaluative variables (such as customer satisfaction and service quality) are key in marketing scholarship because the field views these as being the ultimate drivers of major “loyalty-related” dependent variables (Zeithaml, Berry and Parasuraman, 1996). Over the past 20 years, the field of relationship marketing literature has examined the nature and connectedness of constructs that lie at the heart of effective relationships between consumers and marketing organizations (Morgan and Hunt, 1994; Bansal, Irving and Taylor, 2004). In this field, we have a good understanding of the basic relationships (on loyalty related variables) of these constructs, but we only have a basic understanding of the conditions under which these relationships hold (Fullerton, 2011).

Some take the view that customer engagement is mostly a behavioral issue (van Doorn et.al, 2010). The behavioural element is that engaged consumers do a number of things in response to their engagement related thoughts and attitudes. These behaviors may include such things as repurchase, positive word of mouth through customer referring, social media and/or blogging (van Doorn et.al, 2010). For these researchers, customer action is key to engagement. Very clearly, this is something that marketing practitioners might value because it would lead to positive outcomes for the organization. But it also raises the question as to how customer engagement differs from customer loyalty, particularly behavioural loyalty.

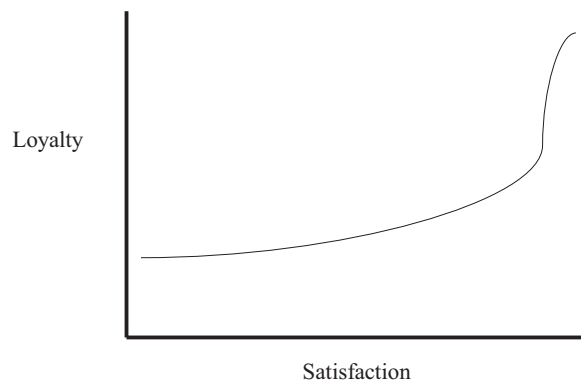
Van Doorn et.al (2010) put forward a relatively complete model of customer engagement that links a number of significant antecedent states to key engagement behaviors to the key outcomes in which a marketing firm is interested. At some level, such a recursive model might be more of a multi-order factor than a conceptual model. Such a holistic view of customer engagement implies that it has significant cognitive, affective and behavioural elements. But it is problematic because everything in the model is conceivably related to (or at least correlated with) everything else. This increases the probability that such a complicated model will be able to explain the behavioural variables of interest. However, it also puts forward a significant, “so what?” Does this complicated model do a better job of explaining outcomes than stand-alone antecedent constructs that are well researched in the discipline? Simply put, does customer engagement really matter?

At a marketing practitioner level, there is an appeal of the customer engagement philosophy. It is obvious that engaged customers will be more loyal and support the company by completing a number of pro-social (from the marketer’s perspective) behaviours. Perhaps it is even easier to understand than variables such as customer delight and commitment. Indeed, from their perspective, the focus on customer engagement might be the natural outgrowth of the relationship marketing era, where marketers have accepted the wisdom of building relationships with their customers (Rust et.al, 2010). Yet, despite the vast quantity of research since Morgan and Hunt’s (1994) landmark piece in the area, we are still uncovering the limits and conditions under which it possible to build effective relationships with customers. Although, it is increasingly clear that customer commitment is the core of effective marketing relationships (Morgan and Hunt, 1994; Bansal, Irving and Taylor, 2004), there are limits to the effectiveness of customer commitment. There are situations where customer commitment is not enough and committed customers still engage in switching behavior or become less loyal over time. At some basic level, marketers have looked to customer engagement to fill this gap. This being said, the confusion over the nature of customer engagement in the broader domain of relationship marketing is somewhat reminiscent of an earlier discussion in the field about the nature of customer delight and it is worthwhile looking at the debate in this area of inquiry to help us understand the potential contribution of notions of customer engagement.

The Nature of Customer Delight

Although researchers have been working with the concept of customer delight for some time, we first started paying attention to it as being something distinct from customer satisfaction about 20 years ago. Formerly, researchers often used “delight” as an anchor in a semantic differential measure for many satisfaction type scales (Spreng and McKoy, 1996). Thus, it is fair to say that it was viewed as an element of an overall affective-type satisfaction judgment. However, the simple, yet important work of Jones and Sasser (1995) put forward that delight might be something different. They argued that, for the purposes of creating true customer loyalty, mere satisfaction was not enough, organizations were required to completely satisfy their customers (Jones and Sasser, 1995). In addition, they found that in many sectors there appeared to be some type of non-linear relationship between customer satisfaction and customer loyalty (Jones and Sasser, 1995). Specifically, the finding was that in competitive market environments, far and away the highest levels of individual customer loyalty existed when the consumer reported that they were completely satisfied. Figure 1. Shows the nature of this satisfaction loyalty relationship.

Figure 1.
Customer Delight in Competitive Markets
Adapted From Jones & Sasser (1996)



Subsequently, the marketing literature has seen a significant discussion of the nature of customer delight and the non-linearity of the satisfaction-loyalty relationship (Oliver et. al, 1997; Rust and Oliver, 2000; Fullerton and Taylor, 2002; Finn, 2012). In

this literature, there are two conflicting positions; the first being that customer delight is a construct and the second being that customer delight is a non-linear relationship that holds under certain conditions. Those who take the view that it is a construct, put forward that delight is an affective state that it is distinctly different from satisfaction, although it may have some common antecedents and some similar effects on dependent variables of interest (Finn, 2012). For others customer delight is a relationship between satisfaction and loyalty and the existence of a non-linear relationship between satisfaction and loyalty provides a simple theoretical explanation for findings which regularly exist in satisfaction research (Jones and Sasser, 1995; Fullerton and Taylor, 2002). The resolution of this debate is very much an on-going process.

The Nature of the Customer Commitment-Loyalty Relationship

However, this debate about the nature of customer delight raises an interesting question for those in the broad areas of relationship marketing and customer engagement. Although customer commitment has been a cornerstone of relationship marketing, we are still seeking to understand the nature and limits of customer commitment. We understand that customer commitment is a complex construct with multiple dimensions (Gruen, Summers and Acito, 2000) and at the same time customers can be committed to multiple targets within a marketing relationship (Jones and Taylor, 2007). We have some appreciation for how the components of customer commitment can both enhance and undermine marketing relationships (Fullerton, 2005). We are also beginning to understand that there are limits to the role that customer commitment plays. In particular, we have found that affective customer commitment is a powerful determinant of customer loyalty-like behaviours (Fullerton, 2005; Gruen, Summers and Acito, 2000; Bansal, Irving and Taylor, 2004). Affective commitment is rooted in identification and attachment and nature of construct is that we (as customers) do business with organizations that we like and to which we are attached (Morgan and Hunt, 1994). Yet, despite its power in explaining loyalty behaviours, it does not explain everything. The recognition of the limits of existing theories of relationship marketing was one of the motivations for the investigation of the nature and effects of customer engagement.

In essence, if customer commitment is a central construct in explaining effective marketing relationships, why do committed customers sometimes defect, leave, switch, change their spending habits in terms of share of wallet? One simple answer to this question is that they are not “engaged” with the company, organization or brand! While others have examined the question as to whether customer engagement is a construct or holistic model (Brodie and Hollebeck, 2011; Spratt et. al, 2009), the question of whether it might have more to do with a non-linear shape of the affective commitment loyalty behavior relationship has yet to be discussed.

There can be no question that the marketing discipline has borrowed extensively from the organization behavior discipline in building our knowledge base on the nature and consequences of customer commitment (Morgan and Hunt, 1994; Fullerton, 2005). Customer commitment is usually viewed as a key mediating construct in that it explains the effect of a number of important background variables [satisfaction, service quality,

trust] on key loyalty related variables [purchase behavior, word-of-mouth, relationship enhancement] (Morgan and Hunt, 1994; Bansal, Irving and Taylor, 2004; Fullerton, 2011).

Despite a significant body of work in the area of customer commitment, the existence of non-linear effects of various forms of commitment has been completely ignored. In the study of employee commitment to the organization, there is some evidence that some components of organizational commitment may have non-linear effects on dependent variables of interest in certain situations (Luchack and Gellatley, 2007). Yet, the reality of customer relationships is that customers who report that they are affectively committed to their relational partner, still leave that partner. One possible explanation for this is that other, previously uninvestigated constructs play a role in explaining the behaviour of consumers in the market place. Those who believe that customer engagement is a construct have proceeded on this path (Spratt et. al, 2009). Others have taken the view that there might be interactions between various forms of commitment that might limit the effectiveness of affective commitment in individual-organization relationships (Fullerton, 2005). This might hold promise for those who believe that customer engagement might be special case where high levels affective commitment produce powerful behaviours.

What Might the Data Show?

Although commitment is a complex, multifaceted construct, most of the time when practitioners talk about the powerful positive impacts of customer commitment, they really mean “Affective Commitment”! As discussed above, affective commitment is rooted a large number of positive cognitive and affective evaluations. It is a powerful predictor of customer behaviours that are important to marketers (Fullerton, 2005). In an increasingly large body of literature on the effect of customer commitment, it always has a linear effect, regardless of the specific behavioral intention examined as the dependent variable. Researchers have always assumed and found linearity in the affective-commitment dependent variable relationship. Yet, an examination of the data over multiple studies (see Fullerton, 2005; Fullerton 2011) demonstrates something a little different, as shown in Figures 2 & 3 below. This might be an indicator of a slight, non-linear effect of affective commitment on two key dependent variables of interest in marketing research in the academic and practitioner communities. However, the big caveat in this analysis is that the non-linear effect does not reach statistical significance! However, it provides a basis for the view that customer engagement might not so much be a holistic model or construct as it is a relationship between product/service attitudes and loyalty related behaviours with respect to the focal product or service.

Figure 2.
Engagement: Affective Commitment → Switching

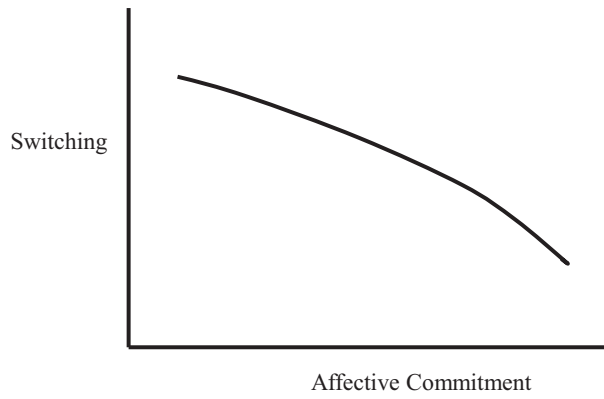
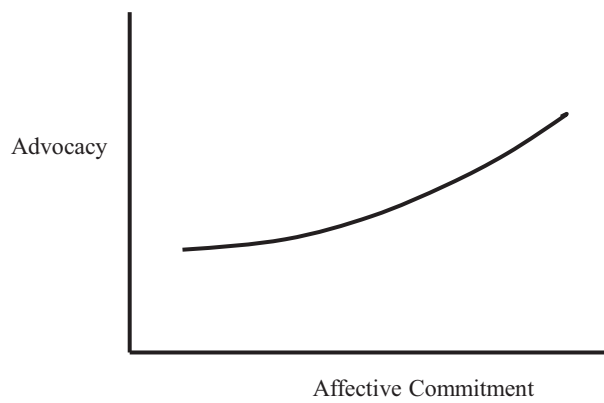


Figure 3.
Engagement: Affective Commitment → Advocacy



Engagement in the Real World of Consumers

Is it realistic that most organizations or brands can have a significant number of engaged consumers? Practitioners have high expectations for customer satisfaction and many organizations expect a large number of consumers (80%+) to report top-box scores on measures of customer satisfaction. Manager and employee incentive compensation is frequently based on this and it indicates of the extent to which the central ideas of the customer delight movement (Jones and Sasser, 1995) have been accepted. Increasingly, organizations also expect that large numbers of customers (80%+) will be willing to recommend their organization/brand to other consumers. This is consistent with ideas that willing to recommend or advocacy represents a connection or attachment on the part of the consumer that is greater than mere satisfaction (Fullerton, 2011). For marketing managers, the key aspect of customer delight is that consumers who report the highest levels of customer satisfaction also have exponentially higher levels of lifetime value. The same logic applies in terms of explaining why so many marketing managers have adopted net-promoter score as a key marketing metric (Reichheld, 2003). Although some question the extent to which the net-promoter score does a good job of explaining firm specific performance results (Keiningham et.al, 2007).

Apple (van Doorn et.al, 2010) and Harley-Davidson (Mollen and Wilson, 2010) are frequently put forward as examples of firms/brands that have significant levels of customer engagement. What is it about these customers and these brands that make them prototypes of the customer engagement phenomenon? The customers are satisfied (indeed delighted) with the products, they love the brands, they often rebuy products over several generations, they are strong advocates for the brand and they may undertake a wide range of “pro-social” behaviors on behalf of the brand! Virtually any organization can have this going on to some degree. But Apple and Harley-Davidson are special cases because of the quantity of their customers who are engaged.

But at the core of these cases lies a simple phenomenon; customers are very satisfied (delighted) with the products and evaluate them very positively, consumers love and identify with the brands and they are prepared do a number of positive things for the organization in response to these thoughts and evaluations. This type of chain is nothing more than Cognition-Affect-Behaviour consistency, which is a relatively old construct in the history of consumer behavior and social psychology (Bagozzi et.al, 1979). The suggestion is that when marketing scholars are conceptualizing customer engagement it exists as nothing more than this among a segment of the customer base. In the case of Apple and Harley-Davidson, it might be a relatively large segment, and in other firms the segment of engaged consumers might be much smaller. This might explain why it is difficult to detect the non-linear relationship between affective commitment and elements of behavioural loyalty. The difficulty arises because of the relatively small numbers of consumers who report the highest levels of commitment and behavioural loyalty in consumer surveys. A phenomenon is tough to detect when it is rare! As many have noted, the data distribution assumptions and distribution requirements might make it somewhat difficult to detect non-linear relationships (Ping, 1995).

Conclusions

For marketing academicians, we might ask whether or not customer engagement is a non-linear relationship between affective commitment and key loyalty behavior variables of interest. Given the discourse about the nature of the customer delight phenomenon, this might be a worthwhile exercise. In settings and sectors where it is known that firms have difficulty creating satisfied, committed and behaviourally loyal customers, it might be required to examine large customer data sets because it will be difficult to detect a slight non-linear relationship.

For marketing practitioners, the challenge of creating engaged customers is to not focus on engagement. Given the recent discussion about customer engagement, for managers and practitioners, it may well be “old wine in a new bottle” (Brodie and Hollebeck, 2011) because the variables and issues are not new. It would be easy to see how one might think of engagement as a repackaging of existing constructs. If there is a way for practitioners to get value from the term, “customer engagement”, it is for them to think of engagement as the state which exists in the market when consumers have high levels of delight, affective commitment, repurchase, customer referencing and other pro-social behaviours. Share of wallet (a key marketing metric in many competitive markets) may well grow exponentially (Kumar et. al, 2010). There is no question that creating that state in the market is significant for organizations and has always been (and continues to be) and continues to be a significant challenge! It is difficult to move customers from being satisfied to being delighted. It is difficult create brand attachment, identification and love. If practitioners can create these states, they have successfully created engagement.

References:

- Bagozzi, R., A. Tybout, S. Craig and B. Sternthal (1979), “The Construct Validity of the Tripartite Classification of Attitudes,” *Journal of Marketing Research*, 16(1), 88-95.
- Bansal, H., G. Irving and S. Taylor (2004), “A Three Component Model of Customer Commitment to Service Providers,” *Journal of the Academy of Marketing Science*, 32(July), 234-250.
- Bolton, R. (2011), “Comment: Customer Engagement: Opportunities and Challenges for Organizations,” *Journal of Service Research*, 14(3), 272-274.
- Brodie, R., L. Hollebeck, B. Juric and A. Ilic (2011), “Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research,” *Journal of Service Research*, 14(3), 252-271.
- Brodie, R. and L. Hollebeck (2011), “Response: Advancing and Consolidating Knowledge About Customer Engagement,” *Journal of Service Research*, 14 (3), 283-284.

- Finn, A (2012) "Customer Delight: Distinct Construct or Zone of Nonlinear Response to Customer Satisfaction?" *Journal of Service Research* 15: 99-110
- Fournier, S. and L. Lee (2009), "Getting Brand Communities Right," *Harvard Business Review*, 87(4), 105-111.
- Fullerton, G. (2011), "Creating Advocates: The Roles of Satisfaction, Trust and Commitment," *Journal of Retailing and Consumer Services*, 18 (1), 92-100.
- Fullerton, G. (2005), "The Service Quality-Loyalty Relationship in Retail Services: Does Commitment Matter," *Journal of Retailing and Consumer Services*, 12(2), 98-111.
- Fullerton, G. and S. Taylor (2002), "Mediating, Interactive and Non-Linear Effects in Service Quality and Satisfaction Research," *Canadian Journal of Administrative Sciences*, 19(2), 124-136.
- Gruen, T., J. Summers and F. Acito (2000), "Relationship Marketing Activities, Commitment and Membership Behaviors in Professional Associations," *Journal of Marketing*, 64 (3), 34-49.
- Jones, T. and S. Taylor (2007), "The Conceptual Domain of Service Loyalty: How Many Dimensions," *Journal of Services Marketing*, 21(1), 36-51.
- Keiningham, T., B. Cooil, T. Andreassen and L. Aksoy (2007), "A Longitudinal Examination of Net Promoter and Firm Revenue Growth," *Journal of Marketing*, 71(3), 39-51
- Kumar, V., L. Aksoy, B. Donkers, R. Venkatesan, T. Wiesel and S. Tillmanns (2010) "Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value," *Journal of Service Research*, 13: 297
- Luchak, A and I. Gellatly (2007), "A Comparison of Linear and Nonlinear Relations Between Organizational Commitment and Work Outcomes," *Journal of Applied Psychology*, 92(3), 786-793.
- Macey, W. and B. Schneider (2008), "The Meaning of Employee Engagement," *Industrial and Organizational Psychology*, 1, 3-30
- Mollen, A., and H. Wilson (2010), "Engagement, telepresence and interactivity in online consumer experience," *Journal of Business Research*, 63, 919-925.
- Morgan, R. and S. Hunt (1994), "The Commitment-Trust Theory of Relationship Marketing," *Journal of Marketing*, 58(3), 20-36.
- Oliver, R., R. Rust and S. Varki (1997), "Customer Delight: Foundations, Findings and Managerial Insight," *Journal of Retailing*, 73(3), 311-336.
- Ping, R. (1995). "A Parsimonious Estimating Technique for Interaction and Quadratic

Latent Variables,” *Journal of Marketing Research*, 32(August), 336-347.

Reichheld, F. (2003), “The One Number You need to Grow,” *Harvard Business Review*, 81 (12), 1-10.

Rust, R., C. Moorman and G. Bhalla (2010), “Rethinking Marketing,” *Harvard Business Review*, 88(1), 94-101.

Rust, R. and R. Oliver (2000), “ Should We Delight the Customer?”. *Journal of the Academy of Marketing Science*, 28(1), 86-94.

Saks, A. (2006),"Antecedents and consequences of employee engagement", *Journal of Managerial Psychology*, Vol. 21 Iss: 7 pp. 600 – 619

Spreng, R. and R. McKoy (1996), “An empirical examination of a model of perceived service quality and satisfaction”, *Journal of Retailing*, 72(2), 201-214.

Sprott, D., S. Czellar and Spangenburg (2009), The Importance of a General Measure of Brand Engagement on Market Behavior: Development and Validation of a Scale,” *Journal of Market Research*, 46(1), 92-104.

Van Doorn, K. Lemon, V. Mittal, S. Nass, D. Pick, P. Pirner and P. Verhoef (2010), “Customer Engagement Behavior: Theoretical Foundations and Research Directions,” *Journal of Service Research*, 13(3), 253-264.

Zeithaml, V., L. Berry and A. Parasuraman (1996), “The Behavioral Consequences of Service Quality,” *Journal of Marketing*, 60(2), 31-46.

ASB 2013
Antigonish, Nova Scotia

Tara Lacey, Faculty
Tourism Management Program
School of Business
Nova Scotia Community College

Donna Sears, Assistant Professor
F.C. Manning School of Business
Director, Atlantic Wine Institute
Acadia University

Benchmarking the Behaviour and Characteristics of Wine Tourists in an Emerging Wine Region¹

Exploratory research was undertaken to benchmark the profiles and preferences of regional wine tourists in Nova Scotia. This paper describes the consumer profiles of three wine tourist segments in the region: Wine Lovers; Wine Interested; and Curious Tourists. The data (n = 780) were collected over an eighteen-month period at wineries, winery events, farmers markets, and on the tour bus of a popular wine tour operator. Results provide evidence supporting the wine tourist typology found in the literature and give insights for tourism bodies as well as individual businesses.

Introduction

The past 25 years have seen steady growth in the production of wine in Nova Scotia (NS) and equally consistent improvement in the quality of the wines produced. There are currently 18 wineries in the province, with at least 2 additional wineries expected to open within the next two years. Ongoing growth is anticipated: the province's wine industry strategy projects an increase of vineyard acreage in the province from 400 to 1000 acres and a corresponding increase in industry revenue from \$7.2 million to \$23 million by 2020 (Winery Association of Nova Scotia, 2009). The success of the industry is demonstrated through national and international attention brought by award-winning wines: 28 awards at the 2009 All Canadian Wine Championships and 14 awards at the 2009 International Tasters Guild Competition. Meanwhile, tourism is a significant driver of the NS economy in general (\$1.3 billion in 2008) and of the Annapolis Valley in particular, with over \$207 million in revenue and 3600 direct jobs in 2008 (Nova Scotia Ministry of Tourism, Culture, and Heritage, 2010). Furthermore, cuisine – including food, dining, wine, and agritourism – is one of NS's core tourism experiences (Nova Scotia Tourism, Culture and Heritage). And there is further evidence of the quality of the Nova Scotia wine tourism

¹ This research was supported by Acadia University Start-up Research funding and a SSHRC Small Institutional Grant.

offering. In 2011, Le Caveau Restaurant (at Domaine de Grand Pré Winery) was named one of the world's top ten winery restaurants by Wine Access.

Nova Scotia is an emerging wine region, and wineries in the province are, by necessity, tourism businesses. As with other new wine regions, the marketing of products focuses on building local residents into loyal customers who then in turn further endorse and augment the region as a destination (Kolyesnikova, Dodd & Duhan 2008). Therefore, an understanding of the region's wine tourists is a key success factor as wineries seek to develop product and experience offerings that will inspire re-purchase (re-visit) intent as well as intent to engage in word of mouth behaviour. Although some wineries in the region were previously making rudimentary estimates of customer characteristics, no systematic research had been done on wine consumers in Nova Scotia prior to this study.

Wine tourism experts have emphasized the need for region-specific wine tourism research, asserting that consumer research, in particular, is not generalizable from one region to another: "...Charters and Ali-Knight (2002) admonished that wine tourist expectations are likely to vary from region to region. No one set of critical success factors will apply everywhere." (Getz and Brown, 2006, 148). An indepth assessment of the market is necessary to create successful marketing communications for any emerging wine region; understanding consumer attitudes contributes significantly to the commercial success of new products (Kolyesnikova, Dodd & Duhan 2008). Moreover, legislation regarding the distribution and sale of wine differs by region, vastly changing the business and competitive environments in which wineries operate. For instance, Nova Scotia's legislation makes it much more lucrative for wineries to sell directly to consumers at the farm gate, versus selling through Nova Scotia Liquor Commission stores. Thus, we cannot assume that wine tourism research conducted in other regions applies to Nova Scotia wine tourists and their desired experiences.

Context

Wine Tourism & Wine Tourists

According to Hall (1996) wine tourism may comprise the visiting of vineyards, wineries, wine festivals and wine shows, where the important motivating factors for visitors are wine tasting and/or experiencing the attributes of a wine region. Charters and Ali-Knight (2002) define the wine tourism experience as encompassing many characteristics including a lifestyle experience, education, links to art, wine and food, tasting and cellar door sales, winery tours, incorporation with the tourism-destination image and a marketing opportunity. Getz and Brown (2006) declare wine tourism as "simultaneously a form of consumer behavior, a strategy by which destinations develop and market wine-related attractions and imagery, and a marketing opportunity for wineries to educate, and to sell their products, directly to consumers." (p. 147). Clearly, wine tourism is a multi-faceted concept, and has not achieved a uniform approach or definition.

Getz and Brown (2006) found three key attributes required for an attractive wine region (from a consumer perspective): core wine product (visitor friendly wineries, knowledgeable winery staff, wine festivals, familiar wineries); essential destination features (attractive scenery,

pleasant climate, moderately-priced accommodation, easy to obtain information, well signposted wine trails); and related cultural experiences (unique accommodation with regional character, fine dining and gourmet restaurants, traditional wine villages).

As Getz (1998 in Charters & Ali-Knight, 2002) explained, there are three perspectives to consider when approaching wine tourism: (1) destination marketing strategy; (2) consumer behaviour; and (3) opportunity for wineries to educate consumers and engage in direct sales. These perspectives are not discrete, therefore, investigation of one perspective informs the others. Wine tourists are also wine consumers in search of (lifestyle) product-related experiences; the motivations of wine tourists are considered an important aspect in the overall understanding of wine regions and wineries of the needs and expectations of their customers; (Alant and Bruwer, 2004; Charters and Ali-Knight, 2002).

The current study addresses an acknowledged gap in the literature by focusing on the consumer behaviour perspective: “Key researchers, in this emerging field of wine tourism, have commented on the lack of published research material available regarding the behaviour and characteristics of the winery visitor or the tourist.” (Charters & Ali-Knight, 2002, p. 311). Despite examining the consumer behaviour perspective, results of this research can be informative for destination marketing and the development of winery product and experience as well. For example, understanding the type of wine tourist, coupled with intention, may direct wineries to the provision of products and services that will satisfy – or delight – customers (Charters & Ali-Knight, 2002). Furthermore, it is acknowledged that cooperation among destination marketing organizations, and others who promote wine tourism – including individual wineries – is a key success factor for wine tourism (Getz & Brown, 2006). Also of importance are the views expressed by destination marketers that every wine label acts as a promotional medium for the region where the wine was produced (Getz and Brown, 2006, p. 149). It is hoped that the results of this research will help to facilitate collaboration.

A comprehensive profile of Nova Scotia wine tourists and potential wine tourists will enable understanding of the key influences in the selection of Nova Scotia wines and visitation to particular Nova Scotia wineries as well as the broader Nova Scotia wine region. By recruiting participants at farm markets as well as wineries, this research overcomes a shortcoming identified in previous wine tourism research: “Most studies of wine tourists have covered only visitors to wineries, not wine consumers in general.” (Getz and Brown, 2006, p. 148). The profile advances the theoretical understanding of hedonic consumption while concurrently offering guidance to tourism operators as they plan and implement hedonic experiences aimed at enhancing customer value. In doing so, the fields of marketing, tourism, entrepreneurship, and rural development are effectively combined.

The current investigation applied a previously developed wine tourism typology that sets out three types: visits of wine connoisseurs and buyers to specific vineyards for buying or scientific purposes (Wine Lovers); visits to vineyards in general, with the aim of understanding process of producing wine and/or wine tasting (Wine Interested) thereby developing long-term customer loyalty; and thirdly visits by way of scenic (wine) routes through wine producing areas linking different vineyards and wineries (Curious Tourists). Through the application of this typology, wine tourists in this emerging wine region can be effectively segmented. Such market

segmentation is significant for wine tourism operators in terms of product development and marketing purposes (Mitchell et al., 2000; Williams & Kelly, 2001) as it provides an understanding of wine tourists and their behaviour.

The Grape and Wine Industry in Nova Scotia

Nova Scotia has a history of grape growing dating back over 400 years. However, commercial wine production – and wine tourism development – is a relatively new phenomenon. The industry has grown from one winery in 1980 to 18 wineries in 2013, with wine sales of \$17.4 million (2011) and wine tourism revenues of \$35 million (2010). Despite its meteoric growth, Nova Scotia's wine industry is marked by small-production wineries and relatively small (less than 10 acre) vineyards (average size is 7 acres). Of the province's 94 farms growing grapes (in 2011) only 10 have more than 10 acres of grapes. In total, the province has approximately 658 acres under vine. Production has grown from 56 tonnes in 1987 to 1264 tonnes in 2011, a 23-fold increase. Thus, the region represents a small – but rapidly growing – share of Canada's \$1.2 billion in revenue and employment related to wine tourism and the grape and wine industry's national \$6.8 billion economic impact (Frank, Rimerman +Co. LLP, 2013).

It is commonly suggested that New- and Old-World wine regions are vastly different and that the consumers may differ by New/Old World as well (Charters & Ali-Knight, 2002). Nova Scotia, of course, is considered a New World wine region. However, the characteristics of this wine region may make Nova Scotia more similar to Old World regions such as southern Europe than new world regions such as Australia or Napa Valley (California). To illustrate: “The structure of the wine industry is substantially different in Europe, with much less concentration of capital, and – particularly in southern Europe – much more small-scale viticulture, often as part of a mixed farming set-up.” (Charters & Ali-Knight, 2002, p. 313). This blend of New and Old World characteristics sets Nova Scotia apart. Thus, it is crucial for region-specific research be undertaken to explicate the characteristics, motivations, and preferences of Nova Scotia wine tourists since they are likely to differ from those in other wine regions.

Method

Wine tourists and potential wine tourists were queried to determine the drivers of behaviour with regard to Nova Scotia wineries. This was accomplished by administering an interview based paper-and-pencil instrument – designed to probe opinions about Nova Scotia wine/wineries, visiting (and re-visiting) intentions, demographic information, and visitor motives – at strategic points throughout the province. Teams of two undergraduate student researchers administered the face to face survey, approximately five minutes in length. Data collections took place at participating Nova Scotia wineries and at farmers' markets, with a focus on the Wolfville Farmers' Market (closest to the wine region) and farmers' markets in the Halifax Regional Municipality (HRM). HRM is the most likely catchment area for day trips to Nova Scotia wineries; it is home to about 384,000 of the province's 938,000 residents and is a reasonable day trip (60-90 minute drive) from the primary wine region (Annapolis Valley).

Participants were asked to describe both wine interest (no interest, limited interest, interested, highly interested) and wine knowledge (no knowledge, limited knowledge, knowledgeable, highly knowledgeable). Consistent with previous research (c.f. Charters & Ali-Knight, 2002; Hall, 1996), interest in wine was used to segment participants because segmentation by self-reported wine knowledge is problematic. Since Nova Scotia is a new wine region, its people are apt to have a lower level of overall wine knowledge than in regions where there has been time to develop a more knowledgeable consumer base. Furthermore, self-measures of knowledge might be skewed by personality characteristics such as self-confidence. As Charters & Ali-Knight explain, "The problem with knowledge is that it is only suggestive of the respondent's motivation as a wine tourist, and is difficult to quantify. Visitors may be knowledgeable without being enthusiastic, and knowledge is probably less useful as an indicator of their activity levels as wine tourists and their relationship with the local tourism market than other psychographic indicators such as interest and attitudes." (Charters & Ali-Knight, 2002, p. 313). Charters and Ali-Knight (2002) further point out the sequential link between interest and motivation, and ultimately behaviour; offering additional support for the use of wine interest as a segmenting variable.

First, descriptive statistics were employed. Then chi-square, z-tests, and ANOVA were used to identify significant differences among the groups.

Results

The results of this research provide an important benchmark for wine tourists in the new wine region and provide evidence that the wine consumer typology (c.f. Hall, 1996; Charters & Ali-Knight, 2002) applies to consumers in the region. It is anticipated that this research will be repeated as the region's wine industry continues to grow and change. Nova Scotia is uniquely positioned as an emergent cool climate wine region whose growth is concurrent with the development of wine tourism as a research stream. Consequently, the current project develops a benchmark that will enable longitudinal study of the development of wine tourism in this new wine region.

Seven hundred eighty (780) people were surveyed at wineries (February; August and September) during special events (Ice Wine Festival; high tourist season; and Fall Wine Festival, respectively) and at Halifax and Wolfville Farmers Markets. The sample represented 40% male and 60% female participants. Most of the participants were well educated, with over 90% having at least some post-secondary education; 27% had more than a Bachelor's degree. Household income was reflective of the educational achievements of participants. Most of the participants (66%) were married or in common-law relationships.

In general, the participants at both farm market and winery locations were interested in wine at some level (98%); only 2% indicated "no interest" in wine. Participants were largely from the Annapolis Valley wine region (37%) or the closest urban area, Halifax Regional Municipality (26%). Other parts of Nova Scotia (6%); the remainder of Atlantic Canada (3%); the rest of Canada (11%); USA (2%); and other International (1%) were also represented.

Participants were segmented primarily by wine interest but wine knowledge (self-assessed) was also recorded. Results of a cross-tabulation on these two variables are found in Table 1. Congruent with expectations – based on both previous research and intuition – those who identified as highly interested in wine were more likely to self-assess as highly knowledgeable than other groups, $\chi^2 (25, 762) = 557.974, p = .000$. To be consistent with previous research which expressed concern around segmentation by self-reported wine knowledge, the variable considered for segmentation was interest in wine. Wine knowledge was considered for one subset (highly interested/highly knowledgeable) to provide comparison to previous studies (c.f., Charters & Ali-Knight, 2002).

Table 1².

Interest in Wine	Knowledge of Wine				Total
	No knowledge	Limited knowledge	Knowledgeable	Highly knowledgeable	
No interest (curious tourist)	6	8	1	2	17
Limited interest (curious tourist)	17	75	3	0	95
Interested (wine interested)	7	314	78	1	400
Highly Interested (wine lover)	0	68	144	30	242
Total	30	465	226	33	754

The typology developed by previous researchers (c.f. Charters & Ali-Knight, 2002; Hall, 1996) is relevant to Nova Scotia wine tourists, despite concerns around generalizability. However, the member characteristics of each of these groups may differ by region. Based upon self-report of wine interest, participants were categorized as follows: wine lovers (31%); wine interested (52%); and curious tourist (12%). Table 2 summarizes the results by wine tourist type. A description of the characteristics of category members follows.

² Valid n=754 of 780 total respondents for wine interest/wine knowledge questions.

Table 2³.

	Wine Lovers	Wine Interested	Curious Tourists
Age			
19 – 24 years	11% _a	14% _{a,b,c}	22% _a
24 – 34 years	21% _b	16% _{a,b,c}	6% _b
35 – 44 years	9% _a	14% _c	9% _{a,b}
45 – 54 years	31% _b	20% _b	26% _a
55 – 64 years	17% _a	25% _{a,c}	23% _a
65 and over	10% _{a,b}	10% _{a,b,c}	13% _a
Gender			
Female	60% _a	61% _a	60% _a
Male	40% _a	39% _a	40% _a
Household Income			
Less than \$25,000	11% _{a,b,c}	12% _a	10% _{a,b,c,d,e}
\$25 000 - \$44 999	6% _{c,d}	11% _a	16% _{d,e}
\$45 000 - \$64 999	15% _{b,d}	19% _a	14% _{a,b,c,d,e}
\$65 000 - \$99 999	19% _{a,b}	17% _a	22% _{c,e}
\$100 000 - \$149 999	22% _a	16% _a	14% _{a,b,c,d,e}
\$150 000 plus	11% _a	8% _a	2% _b
Prefer not to answer	17% _{a,b,c,d}	18% _a	23% _{a,c,d,e}
Highest Level of Education Attained			
Less than Grade 9	0 _{a,b,c}	0 _{a,b}	0 _{a,b,c}
Some high school	0 _{a,b,c}	1% _{a,b}	3% _c
High school graduate	7% _c	5% _b	10% _{b,c}
Some postsecondary	9% _b	14% _{a,b}	25% _c
Postsecondary certificate or diploma	27% _{a,c}	26% _{a,b}	22% _{a,b}
Bachelor's degree	26% _{a,b,c}	28% _a	21% _a
Above bachelor's degree	30% _{a,c}	27% _{a,b}	18% _a
Marital Status			
Single	21% _a	27% _a	32% _a
Married/Common law	70% _b	65% _a	57% _a
Divorced	7% _{a,b}	5% _a	7% _a
Widowed	1% _{a,b}	3% _a	3% _a
Children in Household			
None	82% _a	80% _{a,b}	79% _a
One	7% _a	11% _b	8% _b
Two	8% _a	6% _b	10% _a
Three	2% _a	1% _{a,b}	1% _a
Four or more	0 _a	1% _{a,b}	1% _a

³ Each subscript letter denotes a subset of categories whose proportions do not differ significantly from each other at the .05 level.

	Wine Lovers	Wine Interested	Curious Tourists
Drinking wine for ...			
Less than 1 year	2% _a	2% _a	19% _a
1-5 years	9% _a	18% _b	13% _b
5-10 years	24% _b	19% _c	14% _b
10-20 years	22% _b	20% _c	15% _b
More than 20 years	43% _b	41% _c	39% _b
Typical spending per bottle			
Up to \$20	67% _a	75% _a	80% _a
\$30	30% _b	23% _a	20% _a
\$50	3% _b	1% _a	None (0 _a)
\$100	0 _{ab}	0 _a	None (0 _a)
More than \$100	0	0	None (0 _a)

Wine Lover (31%)

Wine lovers are enthusiasts and buyers who visit specific vineyards for buying or specific purposes. The wine lovers (ie. those who self-identified as highly interested) were experienced drinkers, having enjoyed drinking wine for 20+ (43%) or 10-19 years (47%). This group was well educated, with 30% having more than a Bachelor's Degree. Most were married and travelling with spouse. The most frequent age range was 45-64 (48%) or 25-34 (21%); with the 19-24 and 65+ age groups roughly equal (11% and 10%, respectively). For wine lovers, the wine and winery are sole motivations for the visit, with a strong interest in culture (romance and elegance) as a motivation for visiting.

Previous research suggests gaining wine knowledge is a key driver for this group: "In summary, one could tentatively suggest that the entire 'lifestyle package' is particularly important for the 'Wine Lovers' – more so than the other segments. Food and gaining wine knowledge are key components of their motivation in making the visit." (Charters & Ali-Knight, 2002, p. 315). Wine tourists in NS were equally interested in gaining wine knowledge. Not unexpectedly, in their quest for knowledge, members of this group are the most likely to visit a winery website and are somewhat influenced by expert ratings. Furthermore, tastings play an important role in their winery visits. However, they are least likely to attend an event or festival at a winery; this is likely because they do not need added incentive to visit wineries and prefer to visit at less busy times when they have more opportunities to interact with winemakers/owners/staff. Farmers market booths are a key information source for this group.

Financially, wine lovers are important to wineries because they are willing to spend a bit more on a bottle of wine than the other groups: up to \$20 (67%); \$30 (30%); and \$50 (3%).

A subset of wine lovers who self-identify as highly knowledgeable was recognized by Charters and Ali-Knight (2002). This group was labeled wine connoisseurs: "Whilst the size of this group is small (3.3 percent) ... Given that 60 percent of them are in professional or

managerial occupations, and therefore considered to be high-yield tourists, the attention could well yield substantial benefits to the winery.” (Charters & Ali-Knight, 2002, p. 316). Connoisseurs represent 8% of the participants in the current study. Since this group represents a substantial opportunity for development – and many are apt to influence others (often referred to as market mavens) – they merit further study. Since the key determinant of membership in this category is wine knowledge, educational opportunities can move other wine lovers into this subset.

Wine Interested (52%)

Wine interested visit vineyards with the aim of understanding the process of producing wine and/or wine tasting. Like the wine lovers, the wine interested were experienced wine drinkers, with 41% having 20+ years of wine drinking experience. This is consistent with the dominant age range: 45% are ages 45-64. The other age ranges have roughly equal representation: 15% are 19-24; 16% are 25-34; and 10% are 65 and up. Most of the members of the wine interested category are travelling with a spouse (65%), but they are also the group most likely to be travelling with other family members.

For the wine interested, wine is a pastime rather than a passion, thus they are motivated by other leisure activities. Wine is not the sole focus of travel; local/artisan foods are significant activities. However, learning about wine is also important. Members of the wine interested group are the most likely to attend a festival or event at a winery. People in this category are looking to enjoy wine by the glass, but some indicated that they “rarely drink wine.” Just the same, pairing food with wine is an important purchase influence. Restaurants are an important purchase influence for this group.

Most of the members of this group spend modestly on wine: up to \$20 per bottle (75%), however, since this group comprises over half of wine tourists, they are important to winery success despite their modest spending habits.

Curious Tourist (12%)

Curious tourists happen upon vineyards for reasons other than wine, such as enjoying a scenic route. Many curious tourists have been drinking wine for over 20 years (40%) but others are relatively new wine drinkers: 19% have been drinking wine for less than a year. Although 57% of this group is married, singles were more common in this category than in the others. This group is the most likely to visit a winery with family. For curious tourists, the 45-64 age group dominates (49%); and – consistent with new wine drinkers – a further 22% of the category’s members are 19-24 years of age. The 25-34 age group represents 6% of this category; age 65 and up at 13%. Not surprisingly, this category is the most likely to have at least one child living at home.

This group considers a winery visit a novelty experience; members may attend a festival or event. When visiting a winery, curious tourists are interested in art and architecture. Since opportunity seems to drive these tourists, expert ratings are not important influences of purchase.

Proximity seems important to this group as more than expected live in the local (Annapolis Valley) area. In general, findings for this group mirror those of previous research: “This suggests that merely standing at the cellar door, tasting, has less appeal for them than a more active visit to the winery. Beyond that, and their desire to eat at a winery, it is hard to find any common motivational or educational ground, which distinguishes them.... It is also probably that they would not be wine tourists unless either a wine region was immediately proximate to them, or they happened to be visiting one for more general tourism purposes – a view broadly supported by the work of Hall and Macionis (1998).” (Charters & Ali-Knight, 2002, p. 316). Serendipity, rather than planning, brings the curious tourist to the cellar door.

This group shops for wine on a very modest budget, with 78% spending \$20 or less per bottle; group members do not purchase wines priced over \$30.

Common Characteristics

Despite fitting well into the previously developed wine tourist typology, participants in this study had a number of characteristics in common. There was an overall high opinion of the quality of Nova Scotia wines, indicative of the growth and improvement across the industry. In general, enjoying local artisan foods was more influential in the wine tourism decision than restaurants. Appreciation for the rural setting was likewise an important factor in the decision to visit the wine region and its wineries.

In this study, wine tourists were not obtaining information from winery brochures, visitor guides/visitor centres, travel/consumer shows, travel agents, newspapers or other (ie. non-winery) tourism websites. However, this is more likely to be a factor of where – and in what format – information is available than a reflection of consumer preferences. In the on-going quest to be market driven, future research might profitably be directed at identifying the formats and locations where the region’s wine tourists prefer to acquire information.

None of the groups sought overnight packages as a reason to visit the region. This is consistent with the whereabouts of the sample, since most were local or within a reasonable radius for day trips (ie. HRM). Although this finding might provide discouraging for accommodation providers, it is not unlike the visitor profile in more developed wine regions. For instance, 67% of visitors to Napa, California are day trip visitors (Napa Visitor Industry, 2012 Economic Impact Report). Thus, understanding day trip visitors to wine regions is a key success factor, underscoring the importance of the current study.

Discussion & Managerial Implications

The discussion and managerial implications of this research are intrinsically linked. Therefore, the two are addressed concurrently in the following section. Each of the three wine tourist types is examined in turn, then general conclusions are drawn.

Wine Lover

Wine lovers are interested in tasting and education. Thus, wineries should be sure to provide opportunities for both in order to satisfy this group. Since wine lovers gather information at farmers markets, wineries the purpose of their market booths as advertising and creating awareness. This change of perspective – especially given the influence and lack of price sensitivity of this group – might well encourage wineries to participate in farmers markets, whether or not they see an immediate return on investment on market days. Finally, wine lovers seek reasons to re-visit. Therefore, this group might appreciate offerings such as loyalty programs or wine clubs. Wine clubs might be especially appealing to the wine connoisseur subset, since wine club membership has been related to higher wine knowledge (Getz & Brown, 2006).

Wine Interested

The wine interested group is the most likely to be influenced by festivals and events, thus, the preferences of this group should be kept in mind during the event planning process. Since members of this group are not necessarily wine drinkers, other activities – such as education about wine and how to pair wine with food – should be integrated into events and these opportunities communicated through promotional materials. While in the wine region, this group is interested in pursuing other leisure activities, indulging in local/artisan foods, and eating at restaurants. Therefore, this group may be likely customers at a winery restaurant. Furthermore, development of packages and/or cross-promotion with providers of leisure activities in the region may be a useful strategy for attracting these consumers.

Curious Tourist

Since proximity is important for curious tourists, Nova Scotia wineries should look for members of this group in their own backyards. To illustrate this point, Annapolis Valley residents make up a large proportion of the members of this group in the current research. Clearly, this has implications for promotional efforts, which might profitably be geographically targeted. Curious tourists are often traveling as a millennial and parent mix. Since many (19%) have been drinking wine for less than a year, wineries have the opportunity to convert these new drinkers to loyal customers and help move them through the typology to the wine interested group and on to wine lover group. As the province continues to advance as a wine region, the development of scenic touring routes (i.e Wine Routes) will further capitalize the economic opportunities associated with this group of curious tourists.

Limitations and Future Research

Limitations

As with any study, this research has inherent limitations that must be acknowledged. Principle among them is the choice of venues for data collections. Participants who completed the survey at farmers markets are likely to be a friendly audience with an affinity for local products. Furthermore, participants at a particular winery may be more likely to report positive feelings

toward the brand due to a recent interaction with the wine product and winery experience. Thus, the choice of locations may have had a polarizing influence on the data. This potential limitation can be mitigated by using similar data collection sites during subsequent research. This strategy will ensure that data are comparable across studies.

Future Research

As previously mentioned, this same study is expected to be revisited over time. Thus, longitudinal research can examine the development of wine tourism in this New World wine region over time. The results will be helpful, not only for the grape and wine industry in Nova Scotia, but also for a more thorough theoretical understanding of the organic growth of a wine region. To our knowledge, this sort of research program has not been undertaken elsewhere in the world. Future iterations of the study should strive to verify that the typology is appropriately applied to wine tourists/consumers in the region through the use of cluster analysis, confirmatory factor analysis, and other appropriate techniques for identifying groups.

The development of a NS wine tourist profile – as described herein – forms one part of a three-pronged initiative. The other aspects of the research are: (1) a comprehensive study of NS winery websites and the elements that inspire virtual visitors to become real-life visitors (completed work funded by a Harrison McCain Foundation Emerging Scholar Award); and (2) a thorough investigation of visitors experiences at NS wineries, including desired elements of these encounters that may not currently be integrated. Some research has been conducted to explore motivations and decision making in a theoretical context, however the wine tourism product, or experience, has not yet been modeled (Getz and Brown, 2006). The investigation of visitor experiences at Nova Scotia wineries is currently underway. Student researchers administer the experience measurement instrument to wine tourists on Wolfville's Magic Winery Bus. Passengers will have visited up to four local wineries during the tour.

A more specific investigation of wine tourist motivations would also be informative. As Charters and Ali-Knight explain, "There are two things which the literature of wine tourism in the English-speaking world must encompass if it is to avoid to [sic] being one-dimensional. Wine tourism is rarely a discrete activity, but will probably be undertaken in conjunction with some or all of rural, eco-cultural or adventure tourism, and its participants are unlikely to separate the various tourism forms. Secondly, the wine tourists may come with various cultural paradigms, which might make them approach the processes with different expectations and requirements." (Charters & Ali-Knight, 2002, p. 318). Motivation research could examine both the reasons for choosing a specific winery and the reasons for choosing to visit a wine region.

Finally, it would be beneficial to explore the factors that influence level of wine interest and attendant membership in the wine tourist categories. A deeper understanding of these influences could help destination marketing organizations and wineries to shift consumers into higher involvement – and thus more lucrative – categories. For instance, how a winery might inspire a curious tourist who happens into the winery to become a wine interested or even a wine lover. It is anticipated that both involvement and customer loyalty may play a role. For instance, a customer who becomes loyal to a particular winery may experience higher involvement with the

wine category, spurring higher interest and a thirst for knowledge. More members in the wine lover category should translate to a rosier bottom line for the region's wineries.

Conclusion

This work provides support for the existence of different segments of wine tourists in our region (consistent with previous work by Charters and Ali-Knight (2002) and Hall (1996), among others) and evidence for the broad use of the typology that these authors developed. However, a word of caution is advised. Groups – or even couples – may represent more than one wine tourist type. It is important for tasting room staff and others responsible for the visitor experience to be cognizant that each visitor who comes through the door may have a different set of goals and expectations for the winery visit. Moreover, future research should more systematically investigate the structure of the typology, to ensure that it is appropriately applied to wine tourists in this emerging region.

Across the wine tourist types, buying local was a significant motivator for choosing Nova Scotia wine. This finding reflects the popularity of the locavore movement within the province. Since legislation makes it difficult for Nova Scotia wineries to sell their product outside the province (even in the rest of Canada), this finding is heartening.

There was a high content of wine lovers and wine interested in this study. It is imperative that local wineries meet the needs of these segments by continuing to provide high quality wines and excellent visitor experiences while interest is high. Wineries must avoid alienating the wine lovers and wine interested; they must not risk turning these potential loyal customers into curious tourists.

References

- Alant, Karin & Bruwer, Johan, "Wine Tourism Behaviour in the Context of a Motivational Framework for Wine Regions and Cellar Doors" *Journal of Wine Research*, 15 (1), (2004), 27-37.
- Barber, Nelson, "Wine consumers information search: Gender differences and implications for the hospitality industry," *Tourism and Hospitality Research*, 9 (3), (2009), 250-269.
- Charters, Steve & Ali-Knight, Jane, "Who is the wine tourist?" *Tourism Management*, 23, (2002), 311-319.
- Frank, Rimerman +Co. LLP, "The Economic Impact of the Wine and Grape Industry in Canada 2011," <http://www.wgao.ca/Uploads/Canada%202011%20Wine%20Industry%20Economic%20Impact%20Report%20FINAL%202013-20-13.pdf> [Accessed 29 August 2013]
- Getz, Donald & Brown, Graham, "Critical success factors for wine tourism regions: a demand analysis," *Tourism Management*, 27, (2006), 146-158.
- Hall, C.M., (1996). Wine tourism. In: New Zealand. Higham J. (Ed.), *Proceedings of the Tourism Down Under II: A research conference* (pp. 109-119). University of Otago.
- "Invest in Nova Scotia Wines," <http://www.investinnovascotiawine.ca/> [Accessed 29 August 2013].
- Kolyesnikova, N., Dodd, T. & Duhan, D, "Consumer Attitudes towards Local Wines in an Emerging Region: A Segmentation Approach" *International Journal of Wine Business Research*, 20(4), 321 – 334.
- Mitchell, R., Hall, C.M. & McIntosh, A. (2000). Wine Tourism and Consumer Behaviour. In C.M. Hall, L. Sharples, B. Cambourne and N. Macionis (Eds.) *Wine Tourism Around the World, Development, Management and Markets*, Oxford: Butterworth-Heinemann.
- MITCHELL, R. D., (2006). Influences on post-visit wine purchase (and non-purchase) by New Zealand winery visitors. In: Carlsen, J. and Charters, S (Eds.) *Global Wine Tourism: Research, Management & Marketing* (pp. 95-109). CABI, Wallingford.
- "Napa Visitor Industry, 2012 Economic Impact Report," <https://www.dropbox.com/s/c7mc6yy1coanccz/2012%20Economic%20Impact%20Quick%20Facts.pdf> [Accessed 29 August 2013].
- Nova Scotia Ministry of Tourism, Culture, and Heritage, "2010 Tourism Plan," http://www.gov.ns.ca/tch/pubs/NSTCH_MarketingBook.pdf [Accessed 28 February 2013]

Williams, P. & Kelly, J., "Cultural wine tourists: product development considerations for British Columbia's resident wine tourism market," *International Journal of Wine Marketing*, 13(3), 59-77.