Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2010 Proceedings

International Conference on Information Systems (ICIS)

2010

SIGNALING IN CONTENT SHARING PLATFORMS

Ashley A. Bush
Florida State University, abush@fsu.edu

Amrit Tiwana
University of Georgia, tiwana@uga.edu

SIGNALING IN CONTENT SHARING PLATFORMS

Research-in-Progress

Ashley A. Bush
Florida State University
College of Business
Tallahassee, FL 32306-1110
abush@fsu.edu

Amrit Tiwana
University of Georgia
Terry College of Business
Athens, GA 30602
tiwana@uga.edu

Abstract

Internet-based peer-to-peer (P2P) content sharing platforms have emerged as a widespread mechanism for sharing electronic content using the Internet. A persistent problem with such platforms is the ex ante assessment of content integrity and quality. In this ongoing study, we address this understudied issue. Using a multi-method research design, we identify using a grounded theory building approach three broad classes of signaling mechanisms associated with the content, contributor, and network that users integratively use to assess the risk-benefit tradeoffs in downloading a given unit of content (e.g., a file). We propose that these signals influence users' holistic perception of risk-benefit differential, and in turn influence the likelihood of downloading content files. We describe the status of this research-in-progress study. Our primary expected contribution is a middle-range theory of signaling that predicts how signaling mechanisms influence user behavior in such platforms.

Keywords: Peer-to-peer, signaling, conjoint, platforms

Introduction

Internet-based peer-to-peer (P2P) content sharing platforms have emerged as a widely adopted mechanism for users to share and exchange a variety of electronic content such as music, video, documents, and software. We consider P2P file sharing platforms, a type of peer-to-peer content sharing platform (Androutsellis-Theotokis and Spinellis 2004) as markets for the exchange of digital goods (Shapiro and Varian 1999). An ongoing problem in such networks from an end user perspective is one of assessing the quality of content. Electronic products such as content files share the characteristics of other information products in that they are consumption goods whose quality can only accurately be assessed *ex post*. There are legitimate vulnerabilities that arise from malicious content or files that are pervasive on such networks. Therefore, as in any market, users of such networks possess incomplete or limited information about the underlying characteristics that can be used to assess ex ante the quality of an individual piece of digital content. They must use the limited available information or knowledge that they have to make judgments about content quality and to ultimately to make the decision to download a file. The issue of how end users make such judgments has largely been neglected in both theory development and empirical field studies.

How do users of P2P content sharing platforms then assess the quality and integrity of content, and how do such assessments influence their content downloading decisions? We build on the notion of signaling to develop and test a "middle-range theory" (Van de Ven 2007, p.142) of signaling in P2P markets. We use a grounded theory development approach based on observation of user behavior in 100 P2P platforms which is then used to develop an empirically testable model. The crux of the proposed theory is that users rely on three classes of signals—contributor, content, and network signals. The information conveyed by such signals influences their holistic, integrative assessment of risks versus benefits in downloading a given unit of content (e.g., a file), which in turn influences their likelihood of downloading it. Therefore, risk-benefit perceptions that are shaped by these three classes of signals influence user behavior in such networks. The aforementioned information asymmetry means that users must rely on signals conveyed through the P2P network as a mechanism to mitigate any potential risk from downloading a malicious or low quality file. The primary expected contribution of the ongoing study is introducing the notion of signaling to better understand user behavior in P2P content sharing platforms on the Internet.

This research-in-progress paper proceeds as follows. First, we draw on the notion of signaling to develop our theoretical model and hypotheses. Second, we describe the context of our ongoing study. Third, we discuss our research methodology. Finally, we discuss our preliminary data analysis and expected theoretical contributions.

Theoretical Development

Signaling Theory

Signaling theory has its roots in Michael Spence's work on job market signaling (Spence 1973; Spence 1974; Spence 2002). Spence demonstrated that signals manifest themselves as information flows between a sender of a signal and a receiver of a signal. The ability of these information flows to convey information about some unobservable quality of the sender can directly impact the action of the receiver and ultimately affect overall performance of the market (the job market in Spence's work). Thus, a signal is an indirect indicator of some attribute that is not directly observable and the cost of signaling is negatively correlated with the unobservable attribute (Spence 1973, p. 105). The meaning of a signal is dependent upon the domain in which it is conveyed and interpreted (i.e., a signal in a different domain will have a different meaning).

While signals are used as indicators of underlying quality, signals themselves vary in their degree of reliability (Spence 1973). This variability occurs when the interests of the sender of the signal are not aligned with the receiver of the signal, or when an information asymmetry exists between the sender and receiver of the signal. Spence argues that signals are more likely to be reliable when it is prohibitively costly for the signal to convey false information. These costs manifest themselves in two primary ways. First, the more costly it is to produce a signal, the more reliable it is likely to be. Second, if there is a high cost of punishment if caught conveying a false signal, signals are likely to be more reliable. Hence, high costs act as a mechanism to deter phony signals.

Markets also themselves embody certain characteristics that impact their ability to convey reliable signals (Spence 1973). Markets with large numbers of participants or with individuals who participate infrequently may have

difficulty transmitting reliable signals. In these situations, there is no motivation for individuals to invest in projecting credible signals. P2P platforms embody some of these characteristics, in particular infrequency of interaction between users and content contributors. For example, there might be no expectation of continued direct interaction between a user downloading a specific piece of electronic content and the originator of that content.

Signals in Peer-to-Peer Platforms

Peer-to-peer content sharing networks, one type of P2P content sharing platform, are a market for the exchange of digital goods. They are typically used for one-off content sharings through a network of peers and provide a mechanism for search and transfer of files (Androutsellis-Theotokis and Spinellis 2004). Some examples are ones that exchange software (Bittorrent, Freshmeat), music (Kazaa, Blubster), or video (YouTube). Since these platforms are characterized by decentralized control, mechanisms are necessary to ensure cooperation among and accountability of users (Androutsellis-Theotokis and Spinellis 2004).

Users who visit these networks to locate a digital good must evaluate the underlying integrity of the digital good prior to initiating the download. (We refer to "user" as an individual who is looking for a file to download. We refer to "contributor" as an individual who makes a file available for download through the P2P platform.) Since malicious files (e.g., viruses, trojans, worms) tend to permeate these networks, the ability to evaluate the integrity of the file a priori is a key consideration. Users depend on signals to evaluate the unobservable attributes of the digital good prior to deciding whether or not to download the digital good. These signals assist the user in evaluating the risks versus benefits involved in downloading a file. The ability to make this assessment should impact the user's eventual decision to download the content for subsequent consumption.

To develop our typology of classes of signals in P2P content sharing platforms, we conducted a qualitative, grounded analysis of 100 P2P networks following the approach recommended by Glaser and Straus (Glaser and Strauss 1967). We identified the major properties of these networks: (1) the age of the P2P network, (2) the type of content distributed via the network (e.g., audio, video, software, images, etc.), (3) the types of quality control mechanisms used, and (4) the major problems faced by the users of such platforms. We examined the patterns and thematic commonalties across this wide array of P2P platforms to identify the commonly used signaling mechanisms. Our qualitative analysis led us to identify three broad classes of signals that are widely used: contributor signals, content signals, and network signals. These three classes of signals encompass nine specific types of signals identified through our grounded observational analyses that are summarized and defined in Table 1.

Contributor signals refer to types of information about the originator of a given electronic unit of content such as a file that signals the quality of that file to other users. Content signals refer to information about a specific electronic unit of content itself that signals its quality to other users. Network signals refer to observable characteristics of the P2P platform itself that signals the quality of a file on that network to potential users. We argue that users triangulate information communicated by these three classes of signals into an integrative, holistic assessment about the risk-benefit differential associated with downloading an individual file from the network. This risk-benefit differential in turn influences the user's eventual decision to download a specific file by informing the user's perception of the risks versus benefits associated with the download.

Contributor Signals

The class of contributor signals indicates characteristics of the file contributor that are not directly observable by a P2P platform user. These directly observable signals about unobservable characteristics of the contributor manifest themselves through the contributor's use of the P2P network. Hence, they serve as a reputational mechanism that is created in the context of a specific P2P network and are not easily transferable either in to or out of the P2P network (Bush and Tiwana 2005). Examples of such signals include reputational ratings and rating counts of a given contributor based on the assessments of other users who have downloaded content contributed by a given contributor. Such contributor signals influence users' perceptions of the risks versus the benefits of downloading a file contributed by that contributor. This leads to our first hypothesis.

Hypothesis 1: *Contributor signals in a P2P content sharing platform will influence the user's evaluation of the risk-benefit differential.*

Content Signals

The class of content signals indicates underlying characteristics of a file that are not directly observable by a user. Similar to the way in which warranties serve as signals for consumers under conditions of incomplete information (Boulding and Kirmani 1993), content signals can serve as signals of file quality in P2P networks. Examples of such signals include the net difference between positive and negative ratings that are assigned to a given file by previous downloaders of that file; and the number of previous downloaders who have rated a given file. Such content signals influence users' perceptions of the risks versus the benefits of downloading a specific file. This leads to our second hypothesis.

Hypothesis 2: *Content signals in a P2P content sharing platform will influence the user's evaluation of the risk-benefit differential.*

Network Signals

The class of network signals indicates characteristics of the P2P network that are not directly observable by the user of the P2P file sharing network. The reputation of a firm is often based on consumers' belief in the quality of the firm's products (Shapiro 1983). Hence, signals about the network itself will influence users' perceptions of the risks versus benefits of downloading files. Examples of signals in this class include user confidence that a P2P content sharing platform is not flooded with fake or malicious files (e.g., containing spyware or viruses), presence of contributor identity verification mechanisms, presence of contributor and file reputation systems, and ease of use of such rating systems. Such network signals influence users' perceptions of the risks versus the benefits of downloading a file from a particular P2P content sharing platform. This leads to our third hypothesis.

Hypothesis 3: *Network signals in a P2P content sharing platform will influence the user's evaluation of the risk-benefit differential.*

Risk-Benefit Differential

Risk-benefit differential refers to a user's perceived risks relative to the perceived benefits incurred in downloading a particular file. When users evaluate whether or not to download a file in a P2P network, they use the three classes of signals to assess the risk-benefit differential of downloading a particular file. This integrative perception of risks versus benefits in turn influences the likelihood that they will download a particular file from the P2P content sharing platform. We therefore hypothesize that the effects of the three classes of signals on the likelihood of downloading will be mediated by perceived risk-benefit differential.

Hypothesis 4a: *The perceived risk-benefit differential mediates the relationship between contributor signals and the likelihood of downloading a file in a P2P content sharing platform.*

Hypothesis 4b: *The perceived risk-benefit differential mediates the relationship between content signals and the likelihood of downloading a file in a P2P content sharing platform.*

Hypothesis 4c: *The perceived risk-benefit differential mediates the relationship between network signals and the likelihood of downloading a file in a P2P content sharing platform.*

Figure 1 summarizes the proposed research model.

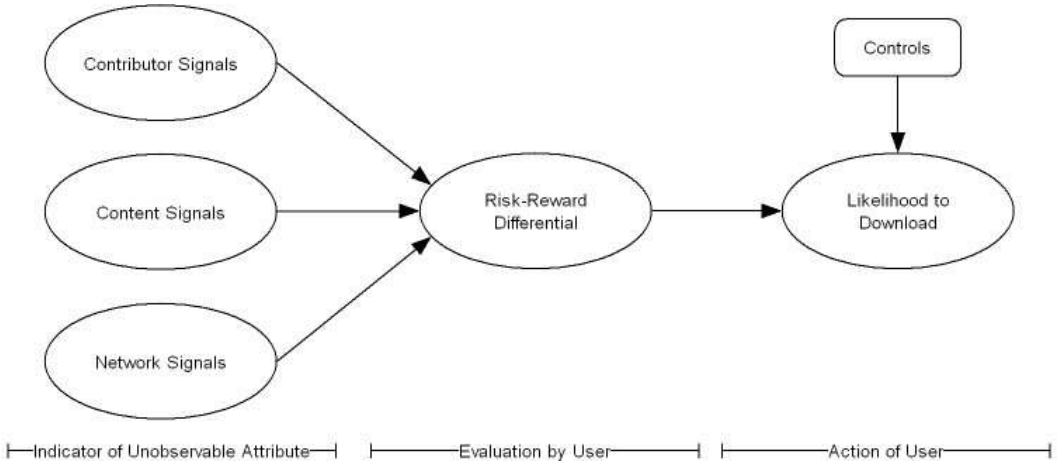

Figure 1. The Proposed Research Model

Research Methodology

Figure 2 summarizes the research methodology used in the ongoing study. The study uses a multi-method, two-phase approach in which the insights from a qualitative phase provide the basis for a subsequent empirical test of the resulting model (Mingers 2001). Phase 1 has been completed and Phase 2 is currently in progress.

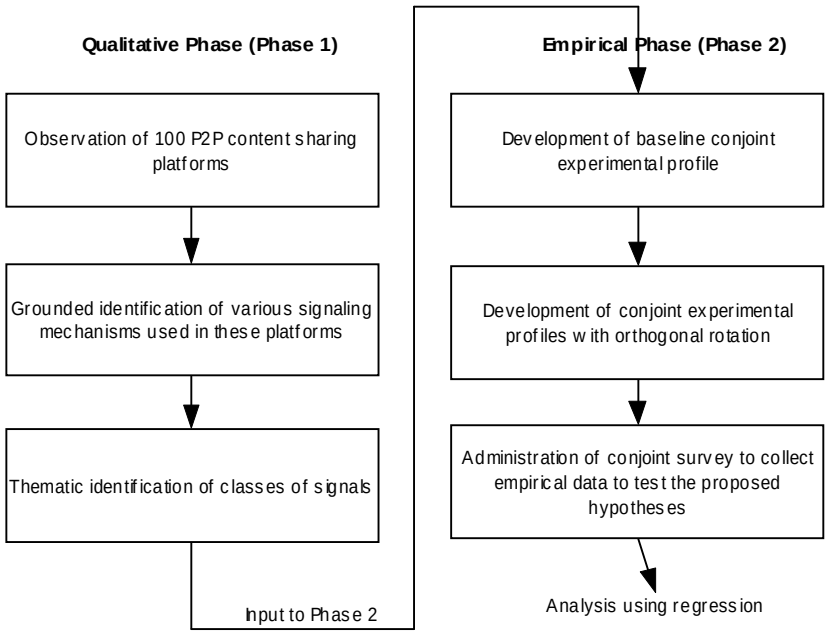

Figure 2. An Overview of the Research Methodology

In the first phase, we observed 100 P2P content sharing platforms. We identified the various signals that were used by each of these platforms. These signals were indirect indicators of platform, user, and contributor attributes that were not directly observable by other users of the platform. We then identified commonalities across these systems and used this to identify thematic patterns of these signals. This led to the insight that such platforms broadly used three classes of signaling mechanisms—content, contributor, and network signals. These together encompass nine different types of signals, which served as the basis for the ongoing Phase 2. The details of the nine specific signals are omitted for brevity and are available from the authors.

To test the hypothesized relationships, we conducted a field experiment using the conjoint research methodology (Louviere 1988). The outcomes from Phase 1 were used to develop a baseline conjoint profile. Conjoint analysis is a multi-attribute judgment analysis technique that is used to determine the combination of attributes that influence a respondent's decision making. Respondents are presented with a number of hypothetical scenarios in the form of conjoint profiles. Each conjoint profile is a different combination of varying levels (high / low) of "conjoint attributes." Our qualitative analysis of 100 P2P networks led us to identify nine different attributes used by the respondent to evaluate two different dependent variables, risk-benefit differential and likelihood of downloading a given file. Table 1 provides a summary of these attributes. Respondents are asked to evaluate these attributes and make a decision about the dependent variable(s) for each profile.

Table 1. Variables in the model in Phase 2 of the study

Signal Class	Signal	Definition
Contributor	Contributor reputation	Rating of the contributor of a given file by other users of that P2P content sharing platform.
	Number of ratings for a given file's contributor	The number of previous downloaders who have rated the contributor of a given file.
Content	File quality rating	The net difference between positive and negative ratings that were assigned to a given file by previous downloaders of that file.
	Number of ratings for a given file	The number of previous downloaders who have rated a given file.
Network	P2P platform trustworthiness	The degree of confidence that the user has that a given P2P content sharing platform is not flooded with fake or malicious files (e.g., containing spyware or viruses).
	Contributor identifiably	The extent to which a contributors' identity can be verified in the P2P content sharing platform. Examples of ways in which this can be accomplished include requiring registration or valid email addresses for creating user accounts on the P2P platform.
	File rating system usage	How extensively a file rating system is used in the P2P content sharing platform.
	Contributor reputation system usage	How extensively a contributor reputation and rating system is used in a P2P content sharing platform.
	Ease of use of ratings	How easy it is to use the file and contributor rating system to make download decisions.

The conjoint methodology is appropriate in situations where respondents may be reluctant to reveal their actual experiences. This methodology overcomes this potential source of bias by presenting respondents with hypothetical scenarios to evaluate. Since P2P networks are often associated with illegal downloads, we chose the conjoint methodology to ensure that respondents did not confound their perceptions of illegal P2P file sharing in responding to the survey. Another consideration in our choice of the conjoint methodology was our focus on multi-attribute decision making. Other research methods such as surveys evaluate a single attribute at a time. The conjoint methodology allows us to simultaneously evaluate the impact of multiple attributes on the dependent variable. Since P2P content sharing networks give users multiple signals to evaluate, the conjoint methodology allowed us to determine the magnitude of the influence of the different signals providing a richer understanding of the user's decision making process than possible with the single attribute approach of a survey.

Development of Conjoint Profiles

We used the orthogonal rotation conjoint algorithm in SPSS to identify the minimal number of conjoint profiles that generate the most information (Green, Helsen and Shandler 1988). In our study, this number was twelve. Each of these twelve conjoint profiles used different combinations of levels for each of the nine signal variables in Table 1. Each conjoint profile therefore described a different hypothetical scenario where the value of each of the nine attributes was assigned a value of "high" or "low". The combination of these nine attributes defines the characteristics of a hypothetical P2P file. Thus, the respondents sequentially evaluate twelve different scenarios.

The contributor signals are indicators of the file contributor that are create within and conveyed through the P2P content sharing network. The contributor signals consist of two attributes: contributor's reputation within the network, and the total number of ratings received by the contributor. The content signals are indicators of the unobservable qualities of the files on the P2P content sharing network. The content signals consist of two attributes: file quality as determined by previous users of the P2P network and total number of file ratings. The network signals are indicators of the P2P content sharing network itself. The network signals consist of five attributes: how extensively the contributor reputation system is used, how extensively the file rating system is used, the user's level of trustworthiness in the P2P network, the extent to which the contributor's identity can be verified within the network, and how easy it is to use the contributor and file rating systems in this P2P network.

There are two dependent variables used by the respondents to evaluate each of the conjoint profiles. First, they are asked to evaluate the risks versus benefits of downloading a file from this P2P network (measured on a 9-point semantic differential scale from "Risks Greatly Exceed Benefits" to "Benefits Greatly Exceed Risks"). Second, they are asked to indicate the likelihood that they would download a file from this P2P network (measured on a 9-point semantic differential scale from "Very Low" to "Very High").

Accounting for Rival Explanations

To account for rival explanations, we also collect data on various control variables including the level of confidence the respondent has in their own evaluations, the approximate number of hours per day that the respondent uses the Internet, how often the respondent uses P2P content sharing platforms, the respondent's prior experience using them, and the respondent's age, educational level, and gender. The first control variable, confidence level of the respondent, is a single item assessment used to determine how much confidence the respondent has in their own evaluation of the conjoint profiles. This variable controls for individual differences in respondent's confidence of their conjoint profile assessments.

Three other extraneous sources of variance that can confound the results are explicitly fixed and invariable. First, use of the hypothetical P2P network is anonymous, downloads can not be traced back to individual users, and the files available for download are completely legal to download. Second, the file available for download is 500 megabytes in size and would take approximately ten minutes to download. Third, the hypothetical platform has been in operation for five years. Collectively, these assumptions remove issues of legality, ethics, user acceptance, and differences in technological capability from the respondent's mental calculus.

Data Collection Approach and Status

The conjoint instrument was pretested with a group of ten individuals who had either technical backgrounds or experience with P2P content sharing networks. They were asked to comment on the clarity of both the instructions and the instrument. Based on the feedback from this pretest group, the instructions and instrument were refined. The conjoint survey instrument was administered to four groups of university students at a major American university. This is an appropriate sampling frame to test the proposed model because such respondents are more likely to be familiar with P2P content sharing platforms and simultaneously exhibit sufficient variance in their use so as to allow rigorous testing of the hypothesized empirical relationships in the model. The response rate breakdown is shown in Table 2 and the overall response rate was 76.15%.

Table 2. Response Rate

Group	Group size	# of Respondents	Response rate
1	60	53	88.33%
2	260	195	75%
3	140	108	77.14%
4	39	24	61.54%
Overall	499	380	76.15%

Preliminary Analysis

The next step in this research-in-progress is to analyze the conjoint data. To test the hypothesized relationships, we use step-wise regression incrementally adding blocks of variables to the model. The control variables are introduced first followed by the contributor signal variables, the content signal variables, and the network signal variables. We then examine the aggregated regression coefficients and statistical significance of each block of variables (representing the three signal classes), and the relative importance of statistically significant predictor in influencing users' decision to download. Additionally, Sobel mediation tests will be used to test the mediation hypotheses (Baron and Kenny 1986).

In our preliminary analysis, we examine the relative importance of each factor in the full model (Table 3). Based upon the beta coefficients, the results suggest that users of P2P networks ascribe the greatest importance to file quality (β=-.29), followed by contributor reputation (β=-.25), followed by P2P platform trustworthiness (β=-.23). These top three signals encompass all there classes of signals—contributor, content, and network. This highlights that users triangulate different classes and sources of diverse signals to make a holistic assessment about the reliability of content in a P2P network. The relatively least importance is ascribed to file rating system usage (β=-.12), ease of use of ratings (β=-.09), and contributor rating system usage (β=-.08). The next step in our analysis is to assess the relative importance of each of the three classes of signals on content consumption decisions in P2P networks and test the mediation hypotheses.

Table 3. Preliminary Analysis

	Contributor Reputation	Content Reputation	Network Reputation
	β(Z-stat)	β(Z-stat)	β(Z-stat)
Constant	(155.93)	(164.87)	(176.13)
Contributor reputation	-.25***(-17.37)	-.25***(-18.37)	-.25***(-19.63)
# of ratings for file's contributor	-.14***(-9.69)	-.14***(-10.24)	-.14***(-10.94)
File quality		-.29***(-21.06)	-.29***(-22.5)
# of ratings for file		-.13***(-9.3)	-.13***(-9.94)
P2P platform trustworthiness			-.23***(-18.16)
Contributor identifiably			-.15***(-11.52)
Contributor rating system usage			-.08***(-6.61)
File rating system usage			-.12***(-9.24)
Ease of use of ratings			-.09***(-6.74)
Perceived risk-benefit diff			
R^2_{adj}(F value)	8.1%***(197.9)	17.8%***(243.1)	28%***(194.1)
ΔR^2(F)	—	9.7%***(265.1)	10.2%***(127.4)

***p ≤ 0.001; **p ≤ 0.01; *p ≤ 0.05.

Expected Theoretical Contributions

Our primary expected contribution lies in proposing and empirically testing middle-range theory of signaling in P2P content sharing platforms. We expect our results to shed light on how users interpret various proposed classes of signals conveyed through P2P content sharing platforms. By identifying and understanding the relative importance of these classes of signals, we hope to gain insight into how such platforms can be designed to mitigate problems associated with malicious content and content of uncertain quality. P2P networks designed to effectively signal

content integrity to users will be moving towards resolving the user-contributor information asymmetry that is inherent in these systems and ensuring the long term sustainability of such platforms.

References

Androutsellis-Theotokis, S. and D. Spinellis. 2004. "A Survey of Peer-to-Peer Content Distribution Technologies," *ACM Computing Surveys* 36(4), pp. 335-371.

Baron, R. and D. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology* 51, pp. 1173-1182.

Boulding, W. and A. Kirmani. 1993. "A Consumer-Side Experimental Examination of Signaling Theory: Do Consumers Perceive Warranties as Signals of Quality?" *Journal of Consumer Research* 20(1), pp. 111-123.

Bush, A. A. and A. Tiwana. 2005. "Designing Sticky Knowledge Networks," *Communications of the ACM* 48(5), pp. 66-71.

Glaser, B. and A. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine Publishing Company, Chicago.

Green, P. E., K. Helsen, et al. 1988. "Conjoint Internal Validity under Alternative Profile Presentations," *Journal of Consumer Research* 15(3),, pp. 392-397.

Louviere, J. 1988. *Analyzing Decision Making: Metric Conjoint Analysis*, Sage, Beverly Hills, CA.

Mingers, J. 2001. "Combining IS Research Methods: Towards a Pluralist Methodology," *Information Systems Research* 12(3), pp. 240-259.

Shapiro, C. 1983. "Premiums for High Quality Products as Returns to Reputations," *The Quarterly Journal of Economics* 98(4), pp. 659-680.

Shapiro, C. and H. R. Varian. 1999. *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston.

Spence, M. A. 1973. "Job Market Signaling," *Quarterly Journal of Economics* 87(3), pp. 355-374.

Spence, M. A. 1974. *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*, Harvard University Press, Cambridge.

Spence, M. A. 2002 "Signaling in Retrospect and the Informational Structure of Markets," *American Economic Review* 92(3), pp. 434-459.

Van de Ven, A. 2007. *Engaged Scholarship: A Guide for Organizational and Social Research*, Oxford University Press Oxford.