INTENTIONS AND SOCIAL INTERACTIONS

J. ATSU AMEGASHIE

INTENTIONS AND SOCIAL INTERACTIONS

Abstract

In psychological games, higher-order beliefs, emotions, and motives - in addition to actions - affect players' payoffs. Suppose you are tolerated as opposed to being genuinely accepted by your peers and "friends". In particular, suppose you are invited to a party, movie, dinner, etc not because your company is desired but because the inviter would feel guilty if she did not invite you. In all of these cases, it is conceivable that the *intention* behind the action will matter and hence will affect your payoffs. I model intentions in a dynamic psychological game under incomplete information. I find a complex social interaction in this game. In particular, a player may stick to a strategy of accepting every invitation with the goal of discouraging insincere invitations. This may lead one to *erroneously* infer that this player is eagerly waiting for an invitation, when indeed his behavior is driven more by strategic considerations than by an excessive desire for social acceptance. I discuss how being tolerated but not being truly accepted can explain the rejection of mutually beneficial trades, the choice of identity, social exclusion, marital divorce, and its implication for political correctness and affirmative action.

JEL Code: C73, J16, Z13.

Keywords: guilt, intentions, psychological game, second-order beliefs, social interaction.

J. Atsu Amegashie
Department of Economics
University of Guelph
Guelph, Ontario
Canada N1G 2W1
jamegash@uoguelph.ca

1. Introduction

In standard economics and game theory, only actions should affect payoffs. Intentions are irrelevant. It is the final outcome that matters not the process. But there are clearly situations where intentions affect payoffs. The same action might induce different payoffs depending on the intentions of the parties or players. This implies that beliefs about beliefs (i.e., intentions) affect players' payoffs. Indeed, these are games in which a player's emotions like surprise, joy, and disappointment affect her payoffs. As Geanakoplos et al. (1989, p. 61) observe "[a] player's emotional reactions cannot in general be independent of his expectations and his interpretation of what he learns in the play of the game". They refer to such games as psychological games and provide the first characterization of equilibria. As Kolpin (1992, p. 218) notes in these games "… no single utility function characterizes a player's preferences over the physical outcome set."

Geanakoplos et al. (1989, p. 69) present the following example:

"… a man is deciding whether or not to give a woman flowers. … The woman may be unhappy for either two reasons: she might expect flowers and not receive them, or she might conclude from his behavior that he is willing to disappoint her. Thus, even if she is not expecting flowers but believes that he thinks she is expecting flowers, she will be unhappy not to receive flowers, because this indicates his willingness to disappoint her."

Gilboa and Schmeidler (1988) also analyze a related class of games. Consider the following example in Gilboa and Schmeidler (1988, p. 216):

"Mr. A hates Mr. B for 'wrongs' B did to him. Assuming Mr. A can follow a strategy which will hurt B, A will choose such a strategy for the sake of revenge. However, the revenge will be sweeter if A knows that B anticipates it. That is to say, the same play of the game may result in a higher payoff for A if B *knows* that this indeed is going to be the play of the game."

Clearly, the above examples show that beliefs and intentions affect payoffs. Falk, Fehr, and Fischbacher (2000), Brandts and Sola (2001), Falk et al. (2003), McCabe et al.

(2003), and Offerman (2002) present experimental evidence which support the idea that intentions matter in reciprocal relationships. Indeed, intentions matter in other important ways. It is the basis for the legal distinction between murder and manslaughter and partly explains the attitudes of certain groups towards racial profiling.

In a very important contribution, extension, and application of the theoretical work of Geanakoplos et al. (1989), Rabin (1993) modeled the idea that people reward actions which have kind intentions and punish actions which have unkind intentions.[1] This has been recently extended in Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006). Dufwenberg and Kirchsteiger (2000) have applied psychological game theory to study wage under-cutting. Also, Ruffle (1999) studies a psychological gift-giving game and Dufwenberg (2002) studies how feelings of guilt and beliefs affect the payoffs of spouses and sustain investments in marriage. In a related contribution, Charness and Dufwenberg (2005) theoretically and experimentally show how guilt aversion can cause people to fulfill their promises in partnerships.

The purpose of this note is to analyze a related class of social interactions. Suppose you are tolerated as opposed to being genuinely accepted by your peers and "friends". In particular, suppose you are invited to a party, movie, dinner, etc not because your company is desired but because the inviter would feel guilty if she did not invite you, or you got a job at an elite institution but you wouldn't have been offered the job if you were not a minority, or someone gives you a present because they felt obliged to do so not because they really wanted to give you a present? Or a friend is expected to give

---

[1] Note that Gilboa and Schmeidler (1988) differs from these papers in one key respect. Their main goal is to point out that common knowledge and "common sense" are inconsistent in games where beliefs or information about the outcome affect payoffs (see their proposition 2). Their work may be interpreted as drawing attention to the limitations of traditional game theory in analyzing psychological-type games.

you a phone call because you need her emotional support. If you have a caller ID and you think she is making the call reluctantly, will you answer the phone? If your boss, supervisor, or professor tells you to feel free to come talk to her anytime you encounter problems in your work, will you take her up on that offer, if you thought she was making the offer grudgingly? Does one's enjoyment from sex depend on whether her partner's intention is a long-term relationship or casual relationship? Will the answer affect the decision to accept or reject an invitation into a sexual relationship? In all of these cases, it is conceivable that the intention behind the action will matter and hence will affect your payoffs. The average reader may be able to relate to some of these situations from personal experience. These examples are common and interesting social interactions worthy of study. They are the basis of friendships and relationships at work, school, church, and in our daily lives. They determine who we choose to go to lunch with, play with, and in general socialize with. They determine the frequency and enjoyment of our social interactions.

It is important to note that I assume that there is already some kind of superficial, implicit, or lower-level relationship between the two parties. For example, they may work at the same place. The question is "will the parties take the relationship to the next level in a world where the intentions behind actions matter?" And since it takes two to tango, one party must have the option of inviting or not inviting the other party and the other party must have the option of accepting or rejecting the invitation.

To the best of my knowledge, this is the only paper, apart from Battigalli and Dufwenberg (2005), that examines a dynamic psychological game under incomplete information. However, my analysis is restricted to a specific albeit important context.

Battigalli and Dufwenberg (2005) present a more general analysis of dynamic psychological games. But they do not examine the kind of social interaction that I focus on or the applications herein. The bulk of Battigalli and Dufwenberg's (2005) analyses is restricted to dynamic psychological games with complete information. They briefly explain how their analyses and framework and be adapted to examine dynamic psychological games under incomplete information but caution that such games require more structure in specific situations. In their concluding remarks, they note that "… there are a variety of interesting psychological phenomena waiting to be analytically explored." In this paper, I examine a specific and interesting dynamic psychological game under incomplete information.

In the next section, I present a dynamic psychological game model of social interactions under incomplete information. I discuss applications in section 3. I conclude the paper in section 4.

2. A Dynamic Psychological Game of Social Interaction

I focus on the following specific example. Consider two people, 1 and 2. Person 1 has the option of inviting person 2 to a movie, party, dinner or any social event. If person 2 believes that 1 genuinely wants her company, she gets a utility, $v \geq 0$, if she is invited and attends the social event. If she does not feel wanted at the social event, she incurs a psychic cost of $\theta > 0$. It is *important* to note that person 1 need not show that she dislikes 2's company. All that is required is that person 2 believes that person 1 probably does not like her company but is only pretending that she (i.e., player 1) likes her company. So it is player 2's inference about player 1's intentions that matters. Therefore, the *same* action

(i.e., invitation) by player 1 could give player 2 *different* payoffs depending on her beliefs about player 1's intentions. Note that if $v > 0$, person 2 would accept an invitation from person 1 if she did not care about 1's intentions.

Let v be a random variable that is commonly known to be continuously distributed on $[\underline{v}, \overline{v}]$ with density f(v) and corresponding distribution function, F(v), $\underline{v} \geq 0$.

Suppose that nature gives person 1 a two-dimensional type: social and guilt types. These types are independently drawn. If person 1 is of social type $w_H > 0$, then she derives a *benefit* of w_H from 2's company. If she is of social type w_L, then she incurs a *cost* of $w_L > 0$, when she is in 2's company. Let the probability distribution of these types be such that $Pr(w_H) = p$ and $Pr(w_L) = 1-p$, $p \in [0,1]$. Furthermore, person 1 feels guilty, if she does not invite person 2. This imposes a high or low cost of guilt on her (i.e., person 1). Let the cost be $g_H > g_L > 0$, which are distributed as follows: $Pr(g_H) = q$ and $Pr(g_L) = 1- q$, $q \in [0,1]$. Therefore g_H and g_L are her guilt types. Assume that $g_L < w_L \leq g_H$. The players have common priors. All this information is common knowledge. However, person 1's type on either dimension is her private information.

After observing her two-dimensional type, person 1 has two strategies: invite (I) or do not invite (D). Person 2 has two strategies: accept (A) or reject (R) an invitation from 1. The game is sequential. Player 1 is the first-mover and player 2 is the second-mover.

If person 1's social type is w_L, then she really does not want person 2 in her company. And therefore if person 2 accepted an invitation and became aware of this fact, her payoff will be $-\theta$. So if person 1 invites person 2, what belief should person 2 hold

about person 1's type or intention? This matters to person 2, because she wants to know if person 1 is inviting her because she really likes her or because she feels very guilty otherwise. That is, person 2 needs to know whether she is being *tolerated* or being *genuinely accepted*. Person 2 wants to know the motive or intention behind 1's invitation.

Note that player 1 does *not* feel guilty so long as she invites player 2, even if she does not want player 2 to accept her invitation. For example, if her type pair is $\{g_L, w_L\}$, she might invite player 2 and if player 2 rejects it, then she suffers no guilt. While the motivation for this behavior may be straightforward, it may be helpful to elaborate further. One explanation is that player 1 anticipates if she does not invite player 2, then player 2 will say things or act in a manner to make player 1 feel guilty. But if player 1 invites her and she rejects it, then player 2 lacks the moral basis to make player 1 feel guilty or player 1 can justify her behavior on the basis that she invited player 2 after all. Of course, player 1's guilt need not depend on player 2's words or actions.

In this game, player 1 will have to form beliefs about whether player 2 will accept or reject her offer and player 2 will have to form beliefs about player 1's beliefs or intentions (i.e., second-order beliefs). This gives a psychological game in the sense of Geanakoplos et al. (1989) and Rabin (1993). Player 2 will now have to interpret an invitation under the following possible scenarios: Is player 1 inviting her because (i) her social type is w_H, or (ii) her type pair is $\{w_L, g_H\}$, or (iii) her type pair is $\{w_L, g_L\}$ but has nevertheless invited her, hoping that she will reject it and not feel guilty as a result?

2.1 Solving the psychological game

I look for a psychological equilibrium to this game.[2] Note that if player 1 plays D, then player 2 does not have to respond. So the analysis is restricted to player 2's response when player 1 plays I.

Let $\sigma \in [0,1]$ be the probability that player 2 rejects an invitation from player 1 and let $\sigma' \in [0,1]$ be player 1's expectation of σ. Let $\sigma'' \in [0,1]$ be player 2's expectation (second-order beliefs) of σ'.[3]

Notice that if $\bar{v}$ is sufficiently small and θ is sufficiently large, then given that $\sigma'' \in [0,1]$, player 2 will reject an invitation from player 1. In what follows, I do not focus on such extreme conditions unless otherwise stated.

In this psychological game, player 1's optimal strategy is as follows: (i) if her social type is w_H, she will invite player 2; (ii) If her type pair is $\{w_L, g_H\}$, she will invite player 2; and (iii) if her type pair is $\{w_L, g_L\}$ and $-(1-\sigma')w_L = -g_L$ (i.e., $\sigma' = 1 - g_L/w_L$), then she is indifferent between inviting player 2 and not inviting her.[4] So if her type pair is $\{g_L, w_L\}$, then it is an equilibrium response to invite player 2 with probability σ'.

Note that since $g_H \geq w_L$, it follows that $g_H > (1-\sigma')w_L$ necessarily holds since $\sigma' \in [0,1]$. So if player 1's type pair is $\{g_H, w_L\}$, she will invite player 2, *regardless* of player 2's rejection probability. Hence the strategy in case (ii) is optimal. The only

---

[2] Kolpin (1992) argues that traditional game theory can handle psychological games by allowing players to choose their beliefs. However, to the best of my knowledge, this approach has not been adopted in the literature. In non-game theoretic settings, Akerlof and Dickens (1982) and Brunnermeier and Parker (2005) allow agents to choose their beliefs.

[3] As in Rabin (1993), Ruffle (1999) and Dufwenberg (2002), I do not consider beliefs beyond second-order beliefs.

[4] I assume that this condition holds in equilibrium. I shall relax it shortly.

scenario in which player 2's rejection probability could *possibly* influence player 1's invitation decision occurs when player 1's guilt type is g_L.[5] Therefore, to make the analysis interesting, I assume that $\Pr(g_L) = 1 - q > 0$ (i.e., $q < 1$).

I first construct an equilibrium such that $\sigma' = 1 - g_L/w_L$. I shall relax this restriction shortly.

Given player 1's strategy, player 2 computes the posterior probability

$$\rho(w_L|I) = \frac{\rho(I|w_L)\Pr(w_L)}{\sum\limits_{i=L,H} \rho(I|w_i)\Pr(w_i)} = \frac{[q + (1-q)\sigma''](1-p)}{[q + (1-q)\sigma''](1-p) + p}. \text{ Also,}$$

$$\rho(w_H|I) = \frac{p}{[q + (1-q)\sigma''](1-p) + p}.$$

Then player 2 rejects an invitation from player 1 if

$$U(A|I) = \rho(w_H|I)v - \rho(w_L|I)\theta < 0 \tag{1}$$

Since the posterior probabilities depend on σ'', it follows that player 2's payoff in equation (1) depends directly on σ''. This is not the case in standard game theory and it is in this sense that this game is a psychological game.[6] Note also that while Geanakoplos et al. (1989), Rabin (1993), Ruffle (1999) and Dufwenberg (2002) incorporate beliefs directly into the utility function, the dependence of player 2's payoff on beliefs arises naturally in my model via Bayesian updating. Also, player 1's payoff depends on σ'.

Given the condition in (1), player 2 of type $\hat{v}(\sigma'') = \dfrac{\rho(w_L|I)\theta}{\rho(w_H|I)}$ is indifferent

between accepting or rejecting an invitation. It follows that

---

[5] Of course, her social type must be w_L as well.
[6] See, for example, the payoffs in Geanakoplos et al. (1989), Rabin (1993) and Dufwenberg (2002).

$$\sigma = \int_{\underline{v}}^{\hat{v}(\sigma'')} f(v)dv = F(\hat{v}(\sigma'')), \qquad (2)$$

As in Geanakoplos et al. (1989), Rabin (1993), and the other papers cited above, mutual consistency of beliefs requires that $\sigma'' = \sigma' = \sigma$.[7] Given that $F(\hat{v}(\sigma''))$ is a continuous function of σ'' mapping the unit interval [0,1] onto itself, it follows from Brouwer's fixed point theorem that there is, at least, one solution (a fixed point) to (2). Unless otherwise indicated, I assume that the solution is unique. Call it σ^*. This gives the following proposition.

Proposition 1: *In the psychological social interaction game, there exists a psychological equilibrium in which player 1 always invites player 2 if her social type is* w_H *or if her type pair is* $\{w_L, g_H\}$. *If her type pair is* $\{w_L, g_L\}$ *and* $\sigma^* = 1 - g_L/w_L$, *then she invites player 2 with probability* σ^*. *Player 2 rejects player 1's invitation with probability* σ^*.

As an example, suppose v is uniformly distributed on [0,1], then we obtain the unique solution

$$\sigma^* = \frac{\theta q(1-p)}{p - \theta(1-q)(1-p)}. \qquad (3)$$

To ensure that $\sigma^* \leq 1$, we require $\theta \leq p/(1-p)$. This is because if θ is sufficiently high, the condition $\hat{v} \leq \bar{v} = 1$ is violated. Since $\theta > 0$, $\theta \leq p/(1-p)$ requires $p > 0$. However, for $p = 0$, it is easy to see that the equilibrium rejection probability is $\sigma^* = 1$. If it is

---

[7] Notice that simultaneously imposing this restriction and $\sigma' = 1 - g_L/w_L$ makes σ' overdetermined. Later, I shall elaborate on this and construct an equilibrium that satisfies mutual consistency of beliefs but does not have this property.

certain that player 1 dislikes player 2 (i.e., $\Pr(w_L) = 1$), then player 2 should reject any invitation from player 1.

Proposition 1 gives an equilibrium of the psychological game but the two restrictions (a) $(1 - \sigma^*)w_L = g_L$ and (b) mutual consistency of beliefs, resulted in an overdetermined system of equations. In what follows, I construct an equilibrium which does not have this property.

Given equation (3), it follows that $\sigma^* = 0$, when $q = 0$. This result implies that if player 1's guilt type is always g_L (i.e., $q = 0$) and her social type is w_L, then she never invites player 2, since player 2 never rejects an invitation in this psychological equilibrium. Therefore, player 1 invites player 2 with probability $\Pr(w_H) = p > 0$ (i.e., when she draws w_H). In this case, beliefs are consistent and this is indeed an equilibrium given $g_L < w_L \leq g_H$. I summarize this analysis in the following proposition:

Proposition 2: *Suppose* v *is uniformly distributed on* [0,1] *and* $q = 0$. *Then there exists a pure-strategy psychological equilibrium in which (i) player 1 always invites player 2, if her social type is* w_H, *and (ii) player 1 never invites player 2, if her social type is* w_L. *Player 2 always accepts player 1's invitation (i.e.,* $\sigma^* = 0$). *In this equilibrium player 1 invites player 2 with probability,* $\Pr(w_H) = p > 0$.

Suppose we relax the requirement of mutual consistency of beliefs such that $\sigma'' = \sigma$ but σ' need not to be equal to σ''. Therefore player 2's belief of σ' need not be correct, since we no longer require $\sigma'' = \sigma'$. Recent models of games without mutual consistency of beliefs can be found in Crawford (2003), Camerer et al. (2004), and Eyster and Rabin (2005). In these games, players choose best-response strategies given false or inconsistent beliefs. Such behavior may stem from bounded rationality on the part of

players.[8] It is therefore not unreasonable to relax the requirement of mutual consistency of beliefs.

Notice that by relaxing the requirement of mutual consistency of beliefs, the issue of an overdetermined system of equations does not arise. In particular, it is an equilibrium for player 2 to reject player 1's invitation with probability σ^* and for player 1 to (a) invite player 2, if her type pair is $\{w_L, g_L\}$ and $g_L > (1-\sigma^*)w_L$, and (b) not to invite player 2, if her type pair is $\{w_L, g_L\}$ and $g_L \leq (1-\sigma^*)w_L$. In this equilibrium, player 1 has consistent beliefs but player 2 does not.

The problem posed by consistency of beliefs in our psychological game may reflect Battigalli and Dufwenberg's (2005, p. 34) warning that "[P]sychological sequential equilibrium requires more care because the extension of the definition of consistency to games of incomplete information is not obvious."

Given the lack of mutual consistency of beliefs and using equation (3) gives $\partial\sigma^*/\partial p < 0$, $\partial\sigma^*/\partial q > 0$ and $\partial\sigma^*/\partial\theta > 0$.

3. Discussion and Applications

If p is positive but very small, the equilibrium in proposition 2 may appear to a third party as though player 2 is overly eager to socialize with player 1 but player 1 is not that eager to socialize with player 2. However, there is a more complicated social interaction at play. The only way that player 2 can deter player 1 from inviting her with the intention of assuaging her guilt is to appear to overly want to be in player 1's company. That way, if player 1 invites her, she (i.e., player 2) has a higher posterior

---

[8] Indeed, this is the justification in these papers for the lack of mutual consistency of beliefs.

belief that she must have indeed drawn w_H. Notice that $\rho(w_H|I)$ is decreasing in $\sigma'' = \sigma^*$.

Since player 1 has the incentive to invite her in the psychological game, even if her social

type is w_L, player 2 accepts every invitation from player 1 when $q = 0$ in order to

discourage insincere invitations.

Note that $\partial\sigma^*/\partial p < 0$ and $\partial\sigma^*/\partial q > 0$ mean that player 2's rejection of 1's

invitation is higher when p is lower or when q is higher. This suggests that two societies

may have *similar* levels of social exclusion for *different* reasons. In one society, this may

be due to the fact that different groups do not like each other (i.e., very low p) and in the

other society, the groups may like each other (i.e., very high p) but are unable to

communicate this fact due to a very high q.[9]

Furthermore, suppose σ^* is not unique. Then two *similar* societies (i.e., the same

p, q, θ, and f(v) function) could have different levels of social exclusion (i.e., different

σ^*). The possibility of multiple equilibria is well known in models of social interaction

like those in Schelling (1978), Bernheim (1994) and Linbeck et al. (1999). The different

equilibria are driven by the beliefs of agents. However, these beliefs are not explicitly

modeled as functions of the agent's payoffs.[10] In these models, it is only the actions of

players which affect payoffs. The same actions by agents cannot induce different payoffs.

Hence they are standard game-theoretic models of social interaction. The kind of social

interaction in this paper is different because it is a psychological social interaction game

in the sense that beliefs about intentions explicitly affect payoffs.

---

[9] In this example, one might think of the inviters as a dominant group and the invitees as a minority group.
[10] Bernheim (1994) is an exception since a player's payoff is a function of the beliefs of *other* players about her actions. Unlike the case in this paper, a player's payoff is not a function of her own beliefs about other players' beliefs. Battigalli and Dufwenberg (2005) argue that the model in Bernheim (1994) can be re-interpreted, subject to some modifications, as a psychological game with incomplete information.

Loury (1994, p. 435) defines a regime of political correctness as "… an equilibrium pattern of expression and inference within a given community where receivers impute undesirable qualities to senders who express themselves in an "incorrect" way and, as a result, senders avoid such expressions." To the extent that politically-correct language is employed to accommodate those that one does not like or cannot identify with, the strategy by player 1 (the sender) of inviting player 2 (the receiver) when her type pair is $\{w_L, g_H\}$ is consistent with political correctness. Indeed, political correctness can sensitize people and increase the probability, q, that they will be high-guilt types. However, it may have the disadvantage that people are more likely to be suspicious of each other's intentions and hence a decrease in social interactions akin to the higher likelihood of rejections obtained above as q increases. A politically-correct equilibrium could be viewed as one in which some w_L types mimick w_H types. Of course, in this equilibrium there will still be a few people who will deviate from the politically-correct equilibrium. This will be the group of senders with the type pair $\{w_L, g_L\}$ who constitute a small proportion, (1-p)(1-q), if q is very close to 1.[11]

So is political correctness a bad thing if it causes people to be suspicious of the intentions of others? Not necessarily. One thing missing from the model is that person 2 does not derive any disutility from not being invited (i.e., from being rejected). If she did, then we could argue that she derives utility from the mere act of being invited even if she intends to reject the offer. Therefore, political correctness need not be a bad thing if people derive utility from politically correct language *per se*. For example, people may derive utility from others restraining their use of racial slurs or derogatory language, even

---

[11] The result that people with sufficiently extreme preferences will deviate from social norms is obtained in Bernheim (1994).

13

if they know that these people harbor such thoughts. If they do not value political correctness *per se*, then it could be welfare reducing.

Notice that the condition in equation (1) holds when θ is sufficiently high. A very high θ may be the characteristic of a person with a very high sense of identity or self-image, which is consistent with why she may derive a high disutility from associating with people who really don't like her. Associating with people who really do not like her imposes a cost on her similar to the cost stemming from a loss of identity in Akerlof and Kranton (2000).[12] If so, the rejection of 1's invitation when θ is very high may be 2's way of choosing her identity by choosing who to associate with, in the sense of Akerlof and Kranton (2000). Consistent with Akerlof and Kranton (2000), my model will predict that women may reject attempts to entice them to traditionally male professions, if they believe that they will only be tolerated but not truly accepted. A difference between my explanation and Akerlof and Kranton (2000) is that identity is an observable characteristic while intention is not. Intention can be inferred but not necessarily observed.

The analysis may also explain why a *high-ability* person may decide to reject a job offer, if she believes that affirmative action influenced the decision.[13] Of course, this effect is more likely to kick in, if the person has equally-attractive or better outside options, which she believes were not influenced by affirmative action.[14] If high-ability

---

[12] For recent economic models of identity, see Darity, Mason, and Stewart (2005) and Fang and Loury (2005).

[13] An article in the New York Times reported that a young and talented Harvard professor claimed that s/he would leave Harvard, if s/he found out that s/he was hired based on affirmative action. While this may not necessarily be credible, it is nevertheless further testimony of the fact that the intentions behind actions affect people's payoffs.

[14] Notice also that if all that the person cared about was actions *not* intentions, then it may be optimal to accept a job from a high-guilt employer in order to exploit their guilt for material gain (e.g., future higher salary).

people are very proud people with a high sense of identity and self confidence then, as argued in the preceding paragraph, they will tend to have a high θ and would not like to associate with people who do not really like them.

As noted in the introduction, Dufwenberg (2002) studies how guilt can sustain investments in marriage. Without feelings of guilt, a husband could leave a marriage (i.e., divorce) after his wife has supported him in acquiring a lucrative professional training. Mindful of this, the wife will not support her husband and the mutually beneficial professional investment will not take place. Dufwenberg (2002) then modifies the model by assuming that the stronger the husband expects that his wife trusts him to stay, the more disutility of guilt he suffers by choosing to divorce. Using psychological game theory, he shows that the husband's belief about the wife's belief could sustain mutually profitable investments in the marriage. However, Dufwenberg (2002) admits that this explanation cannot account for why some marriages end in divorce after investments have taken place. To be sure, there are several reasons why marriages end in divorce and a single model cannot take account of all these reasons. Dufwenberg (2002) notes that incomplete information may be a reason. This is consistent with the simple model in this paper. Incomplete information coupled with the fact that a spouse may derive enormous disutility from realizing that he or she is only being tolerated may lead to a divorce even after investments have taken place. To see this, suppose a man draws g_H and w_L and invites a woman into a marital and investment relationship, after dating for some time. Then the woman will accept the invitation, if $\rho(w_H|I)v - \rho(w_L|I)\theta \geq 0.$ But the marriage may later end in divorce, when it becomes common knowledge that the husband's type is

w_L. So a high feeling of guilt may initially sustain a marriage and subsequent investments in it but eventually it may not be enough to hold it together.

The analysis has been based on the assumption that player 1 incurs no cost if her invitation is rejected. It is conceivable that if her social type is w_H, she might find a rejection embarrassing. The absence of this cost explains why if player 1's social type is w_H, she always invites player 2. However, we sometimes do not invite certain people into closer relationships not because we do not like them. On the contrary, we like them but we are not sure if it is appropriate to invite them. By keeping the relationship at the original lower level, we do not rock the boat. Indeed, a rejection can even push the relationship to a much lower level. For example, imagine how telling a friend that you are romantically interested in them could damage a hitherto platonic and exciting friendship if your proposal is rejected.

While including this cost of rejection or embarrassment to player 1 might alter some of the results, it will not change the equilibrium in proposition 2. Note that including the cost of rejection to player 1 will induce player 2 to moderate her rejection rate in order to encourage player 1 to invite her if player 1's social type is w_H. But in proposition 2, player 2's rejection rate is at the minimum anyway; it is zero even without this cost. Since there is no rejection by player 2 in this equilibrium, player 1 would not worry about her offer being rejected.

4. Conclusion

I have presented a very simple game of intentions in social interactions in which one player (the inviter) can invite another player (the invitee) into a social relationship. The invitee, who cares about the intentions of inviter, can accept or reject the invitation.

This psychological game gave rise to a complex social interaction. In particular, invitee may stick to a strategy of accepting every invitation with the goal of discouraging insincere invitations. This may lead one to *erroneously* infer that this player is eagerly waiting for an invitation from the inviter, when indeed his behavior is driven more by strategic considerations than by an excessive desire for social acceptance.

Unlike the papers noted above, where beliefs directly enter the utility function, the dependence of players' payoffs on beliefs arose naturally in my model via Bayesian updating. My main contribution lies in the type of social interactions examined and the applications discussed which differ from those in the papers noted above. This is also one of the first papers to examine a dynamic psychological game under incomplete information.

The literature on psychological games is small. This partly stems from the fact that psychological games are more difficult to analyze than standard games. However, the tools of psychological game theory allow us to study certain important social phenomena like the nature of reciprocity as in Rabin (1993), Falk and Fischbacher (2006) and multiplicity of other phenomena discussed in Battigalli and Dufwenberg (2005). I have presented another important social phenomenon which appears to lend itself easily to the tools of psychological game theory. This social interaction is interesting and common. To re-iterate, it is the basis of friendships and relationships at work, school, and in our daily

lives. It determines how often we go to lunch with someone, talk to them over the phone, invite them for dinner, play with them, sleep with them, and in general socialize with them. It determines the frequency with which people invite us into their lives and the corresponding frequency with which we accept those invitations. It could also determine how genuine people think our gratitude is, when we express it. I hope that future work will extend the analysis of the effect of intentions on this type of social interaction. For example, the idea in this paper could ultimately lead to a theory of optimal paranoia or optimal mistrust.[15]

---

[15] I thank Zane Spindler for this observation.

References

Akerlof, G., and Dickens, W.T. (1982). The Economic Consequences of Cognitive Dissonance. *American Economic Review* 72: 307-319.

Akerlof, G., and Kranton, R. (2000). Economics and Identity. *Quarterly Journal of Economics* CXV: 715-753.

Battigalli, P., and Dufwenberg, M. (2005). Dynamic Psychological Games. http://www.u.arizona.edu/~martind1/Papers-Documents/dpg.pdf

Bernheim, B.D. (1994). A Theory of Conformity. *Journal of Political Economy* 102: 841-877.

Brandts, J., and Sola, C. (2001). Reference Points and Negative Reciprocity in Simple Sequential Games, *Games and Economic Behavior* 36: 138–157.

Brunnermeier, M.K., and Parker, J.A. (2005). Optimal Expectations. *American Economic Review* 95: 1092-1118.

Camerer, C.F., Ho, T-H., and Chong, J-K (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics* , 861-897.

Charness, G., and Dufwenberg, M. (2005). Promises and partnerships. *Econometrica*, conditionally accepted.

Crawford, V. (2003). Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions. *American Economic Review* 93: 133-149.

Darity, W.A., Mason, P.L., and Stewart, J.B. (2005). The Economics of Identity: The Origin and Persistence of Racial Identity Norms. *Journal of Economic Behavior and Organization*, forthcoming.

Dufwenberg, M. (2002). Marital Investments, Time Consistency, and Emotions. *Journal of Economic Behavior and Organization* 48: 57-69.

Dufwenberg, M., and Kirchsteiger, G. (2000). Reciprocity and Wage Undercutting. *European Economic Review* 44: 1069-1078.

Dufwenberg, M., and Kirchsteiger, G. (2004). A Theory of Reciprocity. *Games and Economic Behavior* 47: 268-298.

Eyster, E., and Rabin, M. (2005). Cursed Equilibrium. *Econometrica* 73:1623-1672.

Falk, A., Fehr, E., and Fischbacher, U. (2000). Testing Theories of Fairness - Intentions Matter. Institute for Empirical Research in Economics, University of Zürich, Working Paper No. 63. http://www.iew.unizh.ch/wp/iewwp063.pdf

Falk, A., Fehr, E., and Fischbacher, U. (2003). On the Nature of Fair Behavior. *Economic Inquiry* 41: 20–26

Falk, A., and Fischbacher, U. (2006). A Theory of Reciprocity. *Games and Economic Behavior* 54: 293-315.

Fang, H., and Loury, G.C. (2005). "Dysfunctional Identities" Can Be Rational. *American Economic Review (P&P)* 94: 104-111.

Geanakoplos, J., Pearce, D., and Stachetti, E. (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior* 1: 60-79.

Gilboa, I., and Schmeidler, D. (1988). Information Dependent Games: Can Common Sense be Common Knowledge? *Economics Letters* 27: 215-221.

Kolpin, V. (1992). Equilibrium Refinement in Psychological Games. *Games and Economic Behavior* 4: 218-231.

Linbeck, A., Nyberg, S., and Wiebull, J.W. (1999). Social Norms and Incentives in the Economic Welfare State. *Quarterly Journal of Economics* 114: 1-35.

Loury, G.C. (1994). Self-Censorship in Public Discourse: A Theory of "Political Correctness" and Related Phenomena. *Rationality and Society* 6:428-461.

McCabe, K., Rigdon, M., and Smith, V. (2003). Positive Reciprocity and Intentions in Trust Games, *Journal of Economic Behavior and Organization.* 52: 267–275.

Offerman, T. (2002). Hurting Hurts More than Helping Helps: The Role of the Self-Serving Bias. *European Economic Review* 46: 1423–1437.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83: 1281-1302.

Ruffle, B. J. (1999). Gift Giving with Emotions. *Journal of Economic Behavior and Organization* 39: 399-420.

Schelling, T.C. (1978). *Micromotives and Macro Behavior*. New York: Norton.

CESifo Working Paper Series

(for full list see www.cesifo-group.de)

_____

1742 Robert S. Chirinko and Debdulal Mallick, The Elasticity of Derived Demand, Factor Substitution and Product Demand: Corrections to Hicks' Formula and Marshall's Four Rules, June 2006

1743 Harry P. Bowen, Haris Munandar and Jean-Marie Viaene, Evidence and Implications of Zipf's Law for Integrated Economies, June 2006

1744 Markku Lanne and Helmut Luetkepohl, Identifying Monetary Policy Shocks via Changes in Volatility, June 2006

1745 Timo Trimborn, Karl-Josef Koch and Thomas M. Steger, Multi-Dimensional Transitional Dynamics: A Simple Numberical Procedure, June 2006

1746 Vivek H. Dehejia and Yiagadeesen Samy, Labor Standards and Economic Integration in the European Union: An Empirical Analysis, June 2006

1747 Carlo Altavilla and Paul De Grauwe, Forecasting and Combining Competing Models of Exchange Rate Determination, June 2006

1748 Olaf Posch and Klaus Waelde, Natural Volatility, Welfare and Taxation, June 2006

1749 Christian Holzner, Volker Meier and Martin Werding, Workfare, Monitoring, and Efficiency Wages, June 2006

1750 Steven Brakman, Harry Garretsen and Charles van Marrewijk, Agglomeration and Aid, June 2006

1751 Robert Fenge and Jakob von Weizsäcker, Mixing Bismarck and Child Pension Systems: An Optimum Taxation Approach, June 2006

1752 Helge Berger and Michael Neugart, Labor Courts, Nomination Bias, and Unemployment in Germany, June 2006

1753 Chris van Klaveren, Bernard van Praag and Henriette Maassen van den Brink, A Collective Household Model of Time Allocation - a Comparison of Native Dutch and Immigrant Households in the Netherlands, June 2006

1754 Marko Koethenbuerger, Ex-Post Redistribution in a Federation: Implications for Corrective Policy, July 2006

1755 Axel Dreher, Jan-Egbert Sturm and Heinrich Ursprung, The Impact of Globalization on the Composition of Government Expenditures: Evidence from Panel Data, July 2006

1756 Richard Schmidtke, Private Provision of a Complementary Public Good, July 2006

1757 J. Atsu Amegashie, Intentions and Social Interactions, July 2006