More Order without More Law: A Theory of Social Norms and Organizational Cultures

Peter H. Huang
Stanford University and The Law School, University of Chicago

Ho-Mou Wu
National Taiwan University and Academia Sinica

This article applies psychological game theory to study the maintenance of social order. It models the control of corruption in principal-supervisor-agent relationships. The models possess multiple equilibria, which correspond to certain social norms and organizational cultures or their absence. The models demonstrate how expectations concerning the likelihood of corrupt behavior can influence decisions to engage in such behavior via the magnitude of endogenous remorse on the part of the actor.

> Law means good order—Aristotle, *Politics*

> The key to success or failure in government lies in the ruler.
> If the inkling line is properly set above,
> the wood will be straightened beneath it.
> It is not that the inkling line does anything in particular to the wood,
> but rather that the disciplining of the wood in following the inkling line makes it so.
> Thus if the ruler is truly upright,
> honest officials will be commissioned and villainous persons will hide themselves;
> but if the ruler is not upright,
> the wicked will get on in the world and loyal subjects will withdraw into retirement.
> (Translation of section 7 of *The Huai Nan Tzu, Book Nine:*
> *The Art of Rulership* from Ames, 1983, p. 182–183.)

1. Introduction

The role of law in creating and maintaining social order has occupied the attention of legal scholars at least since the time of Aristotle. In recent empirical anthropological field studies, Ellickson (1986, 1991) examines the ways

that ranchers in Shasta County, California handle the damages caused by escaped cattle. He finds that law is far less important than is usually thought and that most people resolve disputes not by recourse to formal laws, but instead by reliance on informal rules or social norms. He proceeds to develop a theory of norms that explains how social order can be achieved without law. In a related article, Ellickson (1989a) proposes the hypothesis of welfare-maximizing norms after studying the practices of high-sea whalers in the pre-steamship era. Other examples where social norms rather than explicit laws control behavior are the lawlessness of academic photocopying; relations between residential tenants and their landlords; and Macaulay's (1963) study of contractual norms. Most existing analyses of social norms are based on the rational actor paradigm and emphasize the repetitive aspect of social interaction.

Our objective in this article is to expand the scope of law and economics by incorporating the role of emotions into the study of how social order, however achieved, can be maintained without further law. We are thus able to overcome Ellickson's (1987, 1989b) criticism that the field of law and economics is based on too narrow an interpretation of rationality. A limitation of traditional or nonpsychological game theory is its inability to handle the case of emotional responses to strategic choices, when those responses depend on expectations about strategic behavior. Emotions that are independent of expectations about strategic decisions can and have been handled by nonpsychological game theory. The difference between emotions that depend on expectations about decisions versus those that do not is an important, but subtle one. An example is the disappointment that a person feels upon seeing a bad film. The magnitude of that disappointment might depend on that person's expectations about how good the film was going to be. Another example is how the amount of shame or guilt from engaging in certain behavior can depend on expectations about the social acceptability of such behavior. This last example can be formally analyzed by applying psychological game-theoretic models to study the maintenance of order in societies and organizations.[1] This article formally models the control of corruption in principal-supervisor-agent relationships. Our models possess multiple equilibria, which correspond to certain social norms and organizational cultures or their absence. The models demonstrate how expectations about the likelihood of corrupt behavior can influence decisions to engage in such behavior.

We anticipate our formal game-theoretic analysis by an informal discussion of emotions in models of rational choice. There is ample psychological, anthropological, and sociological evidence of the importance of emotions in decision-making.[2] Experimental studies provide further evidence that emo-

---

1. For the recent development of psychological game theory, see Geanakoplos, Pearce, and Stacchetti (1989) and Huang and Wu (1992). The framework of psychological game theory can capture those emotional responses that depend on rational beliefs about strategic decisions.

2. See, for example, Frank (1988).

tions motivate human behavior in general and organizational behavior in particular.[3] Obviously, emotions might occur in response to strategic decisions made by others, but emotions might not depend on expectations about such decisions. Emotions of this sort can be incorporated into a strategic model in a straightforward manner by altering the payoffs at the terminal nodes of an extensive form game tree. Our focus in this article is on another category of emotions, namely *expectation-dependent emotions*, which explain why the same action can lead to different emotional reactions on the part of a given person at different times. The same person may or may not experience certain emotions in response to a fixed action, with the difference being the result of different expectations about the choice in question.

Section 2 models the role of social norms in the control of corruption in a principal–agent relationship. Section 3 considers organizational cultures in a three-tier hierarchical principal-supervisor-agent model. Section 4 examines more examples and related literature. Section 5 concludes by discussing the strength and implications of the informational assumptions that we make.

2. Psychological Remorse and Social Norms Relating to Corruption

Economic analysis of corruption postulates that rational actors compare the benefits and expected costs to behaving corruptly.[4] Consider the game of "trust" (from Kreps, 1990) depicted in Figure 1, where A is the principal and B is the agent. We denote by p the probability that B chooses to honor A's trust. This game has a unique subgame perfect Nash equilibrium, with $p = 0$—namely, A does not trust and B would abuse A's trust if it were offered.[5]

Many recent approaches to economic organization (see for example, Dasgupta, 1988) are based on applying folk theorems to a repeated version of this game. Folk theorems, so called because their initial authorship in the 1950's is unclear and, since then, they have become part of the folklore of game theory, are results that provide hypotheses under which any feasible, individually rational outcome of a one-shot game can be obtained in equilibrium for a repeated version of the underlying game. Cooter (1994) defines the "new law merchant" to be norms in specialized business communities that are internalized in a version of Kreps' game that he labels the "investment game." Greif (1994) examines how two eleventh-century groups, the Italian Genoese involved in distant overseas trading and the Jewish Maghribi merchants engaged in long-distance, large-scale trading over all the Muslim Mediterranean, developed different cultural beliefs to mitigate the problem of commitment in a merchant–agent version of this one-sided prisoner's dilemma.

---

3. See, for example, Coursey, Hoffman, and Spitzer (1987).

4. See, for example, Becker and Stigler (1974) and Rose-Ackerman (1978).

5. A subgame perfect Nash equilibrium is a Nash equilibrium whose strategies remain a Nash equilibrium in every subgame. See Fudenberg and Tirole (1991) for the precise definition of subgame perfection. See Baird, Gertner, and Picker (1994) for the application of subgame perfection to legal analysis.

Honor p (10, 10)

B

Trust

Betray $1-p$ (-5, 11)
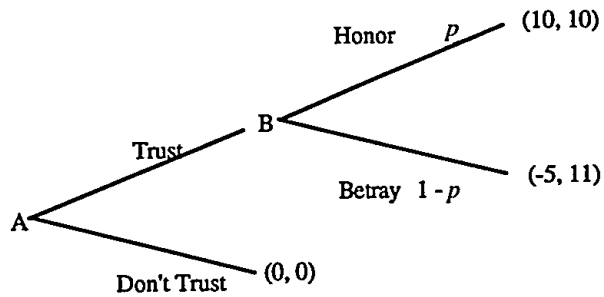
A

Don't Trust (0, 0)

Figure 1. Game of trust.

Repeated games between identified players allows for rewards to encourage cooperative behavior and/or sanctions to punish abusive behavior. They also permit the development of reputations in the presence of asymmetric information. Kreps, Milgrom, Roberts, and Wilson (1982) provide the original model of how a small probability that one is facing a Tit-For-Tat (TFT) player in a finitely repeated prisoner's dilemma can sustain cooperation for all but the last few periods. Recent models of the prisoner's dilemma have a common interpretation that TFT involves a norm of reciprocity. The importance of these formal models cannot be understated because they rigorously provide a role for developing a reputation of being trustworthy. But, as Williamson (1993) and Craswell (1993) point out, such calculative interpretations of honoring trust based on future gains might be missing something important in the meaning of the word "trust."

The traditional economic approach to crime and punishment focuses on anticorruption policies that raise the probability of detection and/or the magnitude of the punishment for corrupt behavior. Typical punishments include fines or monetary losses from termination of employment. But, the "moral costs" of corruption are often also mentioned. Klitgaard (1988, 1991) offers excellent descriptions of how changing attitudes about corruption can be effective as a tool for dealing with corruption. He suggests using education and example, codes of ethics, and organizational cultures to affect the attitudes of agents and clients toward corruption. He deals with exogenous attitudes toward corruption. We suggest that part of the disutility of behaving in a corrupt manner is endogenous and psychological in nature, depending on (rational) expectations about the extent of corruption in one's society or organization. We are able to formalize the intuition in the passage from the *Huai Nan Tzu* quoted at the beginning of this article that the more prevalent corruption is, the less intense is the remorse suffered from corrupt behavior; and conversely, the less corruption there is, the more regret from violating a social norm not to be corrupt.

Suppose the modified game of trust being played is the one depicted in figure 2. There is a single player A facing a continuum of B players. The intensity of remorse that an individual B feels from betraying A's trust depends on B's expectations about A's expectations about the proportion of the

Honor p (10, 10)

B

Trust

Betray $1 - p$ $(-5, 11 - 2\varrho)$

A

Don't Trust (0, 0)
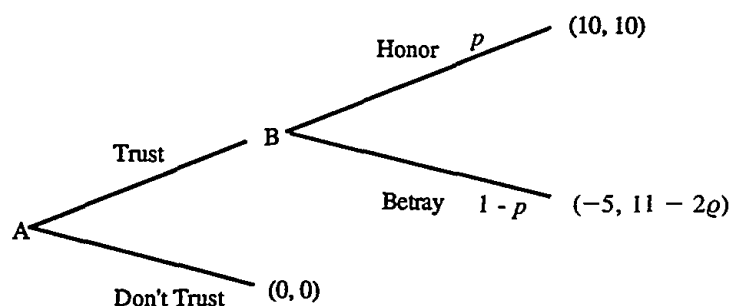
Note: ϱ is B's expectation of A's expectation of p.

Figure 2. Psychological game of trust.

B population who would honor that trust. Let p be the fraction of Bs who choose to honor A's trust; π be A's expectation of p; and ϱ be any B's expectation of π. The psychological remorse that an individual B feels from betraying A's trust depends on ϱ, that individual B's expectations of A's expectations of what proportion of Bs would honor A's trust. Some real-world examples of such interactions are A being the owner of a firm and B being a manager; A being a voter and B being an elected government official; A being a merchant and B being an agent, as in Cooter (1994).

To fulfill the rationality of expectations condition of a psychological equilibrium, p has to equal π and ϱ in equilibrium. There are three psychological equilibria.[6] The first equilibrium has A choosing to trust Bs and $p = \pi = \varrho = 1$, or all of the Bs choosing to honor A's trust. The associated payoffs are (10, 10). A second equilibrium has A choosing never to trust any B and $p = \pi = \varrho = 0$, or all of the Bs choosing to betray A's trust if given the opportunity. The associated payoffs are (0, 0). The third equilibrium has A choosing to trust Bs and $p = \pi = \varrho = \frac{1}{2}$, or half of the Bs choosing to betray A's trust. The associated payoffs are (2.5, 10).

In the first equilibrium, trust occurs and the Bs do not engage in corrupt behavior because they believe that A believes that corruption will not occur. If one B were to engage in corrupt behavior, that B would experience remorse to such a degree that B would rather not engage in corrupt behavior. In the second equilibrium, trust does not occur and Bs would engage in corrupt behavior if given the opportunity, due to their common and rational expectations that corruption is prevalent. Corrupt behavior induces a B to feel remorse only to such a small degree that it makes behaving corruptly dominate behaving honorably. Any individual B would feel little remorse over behaving corruptly because that B believes correctly that every other B would also behave corruptly. In the third equilibrium, A always trusts when half of the Bs engage in corrupt behavior because that still leaves A better off than never

6. A psychological equilibrium requires not only the Nash equilibrium property, but also that players' expectations are correct in equilibrium. See Geanakoplos, Pearce, and Stacchetti (1989) for the precise definition of a psychological equilibrium.

trusting. Half of the Bs engaging in corrupt behavior makes any individual B indifferent between corrupt and noncorrupt behavior.

Our model views the three different equilibrium beliefs as three alternative social norms that are part of three different equilibrium outcomes. Once in equilibrium, no individual is willing to deviate from the equilibrium or the social norm. In contrast to the unique equilibrium for the original game (the one without psychological emotions, shown in Figure 1), the presence of expectation-dependent remorse makes both the first and third equilibria possible. In these two equilibria, the social norms of less corruption are supported by the equilibrium beliefs and psychological emotions. All individuals are actually better off materially as compared to the equilibrium of the original game in Figure 1. Finally, no individual can materially gain by unilaterally becoming unremorseful.

Hwang (1987) discusses three definitions of the Chinese concept of *renqing* (human relations), two of which are relevant to our model. First, *renqing* refers to a set of social norms by which individuals abide in order to live in harmony with others in (Chinese) society. These norms are supported by the desire to preserve or save face (*mianzi*). An interpretation of saving face is to preserve or maintain a reputation for honesty or trustworthiness, calculated for the purpose of having future trading opportunities as in Kreps (1990) or Dasgupta (1988). In Chinese and other Asian cultures, however, losing *mianzi* occurs even if there were no possibility of any future interaction because of the shame that is experienced internally by the individual or by family members. Second, *renqing* denotes the emotional responses that a person has when faced with life's daily events. As Hwang (1987) points out, a Chinese classic, the *Book of Ritual* (*Li-Chi*) states: "What is so-called *renqing?* It consists of happiness, anger, sadness, fear, love, hate, and desire." An individual who knows *renqing* then is one who is empathetic to emotional responses that others experience. In the language of modern game theory, such individuals have common knowledge about the payoffs in a psychological game.

More generally, suppose the game is as depicted in Figure 3. Without loss of generality, we normalize payoffs so that $(0, 0)$ are the status quo payoffs if A does not trust. We denote the "cooperative" payoffs by (c, c) if A trusts and B chooses to honor that trust by not engaging in corrupt behavior. Finally, $-d$ is the sucker payoff equal to the loss A suffers if A trusts B and B betrays that trust; while a is the gain from corrupt behavior, c is the remorse that is expectation-independent, and b reflects how sensitive B's remorse is to B's expectation about A's expectation about the likelihood of corruption.

We assume that the parameters a, b, c, and d are all positive and that $2c < a < 2c + b$ (if $a \le 2c$, then $a - c \le c$, and so in equilibrium $p = 1$ and there is no problem of corruption; while if $a \ge 2c + b$, then $a - c - b \ge c$ and so in equilibrium $p = 0$ and trust is never offered). If $\pi = \varrho = 0$, then $p = 0$ and we refer to corruption as a psychological dominant strategy for B.[7] If $\pi = \varrho = 1$,

---

7. A psychological dominant strategy is defined to be a strategy that if anticipated correctly becomes a dominant strategy.
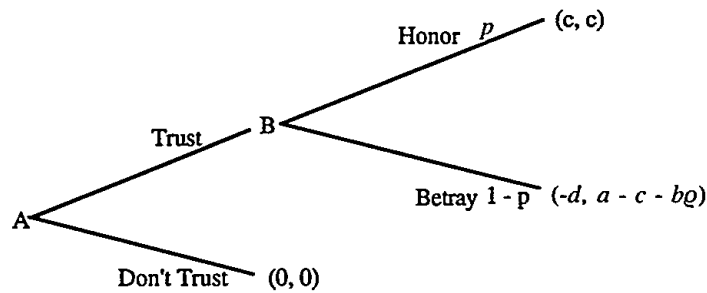
Note: ϱ is B's expectation of A's expectation of p.

Figure 3 .Generalized psychological game of trust.

then $p = 1$ and no corruption is a psychological dominant strategy for B. To find the mixed-strategy psychological equilibrium, we set $a - c - br = c$. This results in the equilibrium $p = \pi = \varrho = (a - 2c)/b$. Notice that because $a < 2c + b$, $(a - 2c) < b$, or $p < 1$. Also, $p > 0$ because $a > 2c$ and $b > 0$. This model leads to some apparently counterintuitive comparative statics results for the mixed-strategy equilibrium—namely, these inequalities: $\partial p/\partial a = \partial p/\partial(a - c) = 1/b > 0$; $\partial p/\partial c = -2/b < 0$; and $\partial p/\partial b = -(a - 2c)/b^2 < 0$. In words, the probability of trust is increasing in the defection payoff and decreasing in the cooperative payoff as well as in the degree of remorse. But, the signs of the comparative statics results become more understandable upon realizing that a higher value of a or $(a - c)$ must be offset by a higher equilibrium $p = \varrho$ and a higher b or c must be offset by a lower equilibrium $p = \varrho$. It is important to remember that these results hold only when comparing mixed-strategy equilibria, as $a - c$ (the expectation-independent net gain to behaving corruptly) or b (the sensitivity of remorse to expectations) change. Finally, we observe that in order for A to choose to trust, $pc + (1 - p)(-d) > 0$, or $p > d/(c + d)$. Thus, the inequality $(a - 2c)/b > d/(c + d)$, or $d < c(a - 2c)/(b - a + 2c) = c(a - c - c)/(c - (a - c - b))$ has to be satisfied for the Bs to have a chance to move in the mixed-strategy equilibrium.[8] An interpretation of this condition is that the absolute value of the "sucker" payoff to A has to be less than the product of the cooperative payoff to B with the ratio formed by dividing the defection payoff to B when $\varrho = 0$ by the difference between the cooperative and defection payoffs to B when $\varrho = 1$.

This section demonstrates how expectation-dependent emotions can act to support social norms against corruption. The payoffs in Figures 2 and 3 capture the intuition that the more prevalent corruption is, the less intense is the psychological remorse suffered from corrupt behavior. It is demonstrated that this gives rise to equilibrium beliefs or social norms that call for less or no

---

8. We note that a mixed-strategy psychological equilibrium is locally stable in the sense that for fixed values of the parameters a, b, c, and d, a small deviation from the interior equilibrium p results in p moving back toward that equilibrium. Such a perturbation, of course, differs from the comparative statics results involving changes in the values of the parameters a, c, $(a - c)$, or b.

corruption. All individuals benefit materially from social norms for behaving honorably. Expectation-dependent emotions can serve the strategic role of sustaining social norms that benefit all of society.

3. Organizational Cultures Related to Corruption

To capture more complex behavior, we extend our basic model to reflect a three-tiered hierarchical organization, as found in Tirole (1986) and Laffont (1990). Klitgaard (1988) defines a similar principal-supervisor-agent structure, which allows for a formal model of a principal choosing not to engage in corrupt actions given rational expectations about how likely agents are not to engage in corrupt behavior and vice versa. Thus, corruption might spread not just among agents but from agents to principals or from principals to agents. In order to prevent a wildfire of contagious corruption, we show that organizational leaders can, by exemplary behavior, set examples for their subordinates and in so doing, cultivate rational expectations that corruption will not be tolerated nor widespread, starting from the top. This formal model supports the passage from the *Huai Nan Tzu* quoted at the beginning of the article.

First, consider the nonpsychological game of a principal-supervisor-agent relationship in a three-tiered hierarchical organization depicted in Figure 4, where A is the principal, B is the supervisor, and C is the agent. Some real-world examples of such interactions are A being a voter, B being a government leader, and C being a government official; A being a shareholder, B being a company's manager, and C being a nonmanagerial employee; or those provided by Tirole (1986). This game has a unique subgame perfect Nash equilibrium: Don't Trust, $p_1 = 0$, and $p_2 = 0$ with (0, 0, 0) as the outcome.

Now, consider a psychological version of this game, depicted in Figure 5. Let p_1 represent the proportion of the B population that chooses not to be corrupt. Let p_2 represent the proportion of the C population that chooses not to be corrupt. Because the B population is made up of identical Bs and the C population is made up of identical Cs, we assume that all Bs have common, symmetric beliefs and that all Cs also possess homogenous beliefs. Let $\pi_1 = $ all Cs' and As' expectations of p_1;. Let $\varrho_1 = $ all Bs' expectations of π_1. Let $\pi_{n_2} = $ all Bs' and As' expectations of p_2. Let $\varrho_2 = $ all Cs' expectations of π_2.
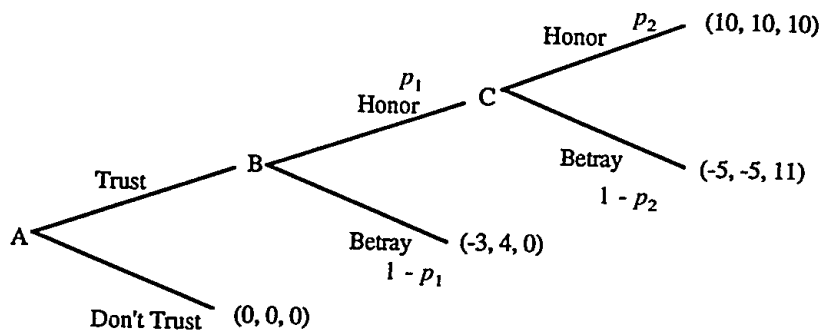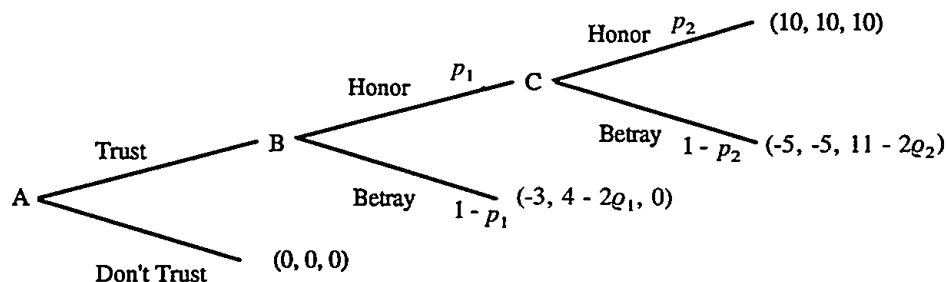

Figure 4. Nonpsychological three-tier game of trust.

Note: ϱ_1 = B's expectation of π_1; π_1 = all C's and A's expectations of

p_1; ϱ_2 = C's expectation of π_2; π_2 = all B's and A's expectations of p_2.

Figure 5. Independent-payoffs case of the psychological three-tier game of trust.

Although we assume that A's expectation of p_1 coincides with all of the Cs' expectations of p_1 and that A's expectation of p_2 is the same as all the Bs' expectations of p_2; this is not crucial. All that matters is that in equilibrium those respective expectations coincide with each other because they all coincide with the underlying endogenous proportions. In the psychological game of Figure 5, any individual B or C experiences psychological remorse that depends on that individual's expectations about others' beliefs over how many B's or C's will choose not to be corrupt. The payoffs for Bs and Cs capture analytically the intuition that the less prevalent corruption is, the stronger is the psychological remorse Bs and Cs feel upon acting corruptly.

One psychological equilibrium is that A always trusts B, $p_1 = \varrho_1 = 1$, and $p_2 = \varrho_2 = 1$. This equilibrium outcome has payoffs of (10, 10, 10). A second psychological equilibrium is that A always trusts B, $p_1 = \varrho_1 = 1$, and $p_2 = \varrho_2 = \frac{1}{2}$. This equilibrium outcome has payoffs of (2.5, 2.5, 10). A third psychological equilibrium is that A always trusts, $p_1 = \varrho_1 = \frac{3}{4}$, and $p_2 = \varrho_2 = \frac{1}{2}$ because in that case Figure 5 reduces to Figure 6, where $\varrho_1 = \frac{3}{4}$ (so that B is indifferent between honorable and corrupt behavior, because $4 - 2\varrho_1 = 2.5$). Individual A trusts because $(0.75)(2.5) - (0.25)(3) > 0$. This equilibrium outcome has payoffs of (1.125, 2.5, 7.5). Finally, notice that once it turns out that C gets to move because B has honored A's trust—whether by deliberate choice to always behave honorably by setting $p_1 = 1$ or by chance when choosing a mixed strategy of $p_1 = \frac{3}{4}$—C is guaranteed a positive payoff. C can commit to honor that is, $p_2 = \varrho_2 = 1$, as in the first equilibrium but C can also choose to behave honorably only half the time as in the second and third equilibria. In both of those cases, A and B still behave in such a way as to ensure that C's payoff is positive. In this sense, then, C can choose to behave "too honestly." A natural question that arises is whether $p_2 = \varrho_2 = 1$ can ever happen even if B chooses $p_1 = \varrho_1 = \frac{3}{4}$. If B believes $p_2 = \varrho_2 = 1$, then B will adopt $p_1 = \varrho_1 = 1$. This is exactly the reverse phenomenon of that described in the passage from the *Huai Nan Tzu* quoted earlier. Because B chooses to act before C does, B can choose $p_1 = \varrho_1 = \frac{3}{4}$ and C can still choose $p_2 = \varrho_2 = 1$,
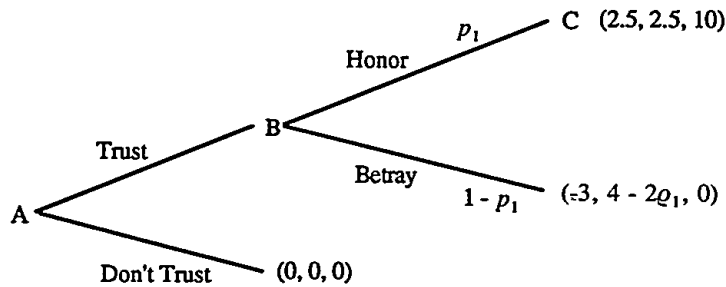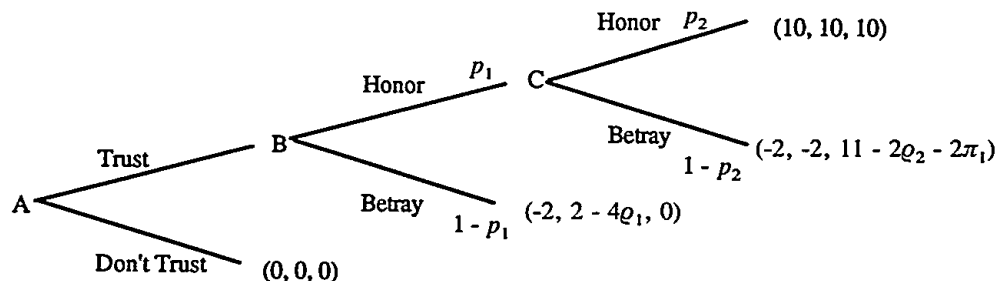
Note: ϱ_1 = B's expectation of π_1; π_1 = all C's and A's expectations of p_1.

Figure 6. A reduced-form version of figure 5.

because C's payoffs are independent of C's expected behavior of the B population. Thus, Figure 5 depicts what is termed the independent-payoff case of the psychological game. With independent payoffs as in Figure 5, C can choose any of the strategies as in the two-player game in Section 2. But B does not know C's action when B decides on a strategy. Hence, we cannot solve the game by applying backward induction.

However, a more plausible case can be made for C's payoff to depend on the expected behavior of both individual Bs and Cs. Consider Figure 7, which depicts the dependent-payoffs case of the psychological game. Individual B, a supervisor or leader of the government, may have an impact on the payoff of C, the agents or government officials, via the actions chosen by B. The payoff of C depends not only on how common corruption is among the C population, but also upon how common it is among the B population (through the proportion q_1). For any C, the expectation-dependent remorse arising from corrupt behavior is more profound if corruption is less widespread among C's peers and C's supervisors. There are now two separate components of endogenous remorse: one depending as before on the expected behavior of one's fellow Cs, the other depending on the expected behavior of one's leader Bs.

If B has decided not to be corrupt but C chooses to be corrupt, then C has

Note: ϱ_1 = B's expectation of π_1; π_1 = all C's and A's expectations of p_1; ϱ_2 = C's expectation of π_2; π_2 = all B's and A's expectations of p_2.

Figure 7. Dependent-payoffs case of the psychological three-tier game of trust.

not only let down other C's expectation of C, but also Bs' expectations of C, because the observation that (10, 10, 10) has not been achieved must be due to corruption by either B or C. This motivates having C's payoffs depend negatively on both ϱ_2 and π_1, where ϱ_2 = C's expectation of others' expectations that C does not behave corruptly and π_1 = C's and A's expectations that B does not behave corruptly. If $p_1 = \varrho_1 = \pi_1 = 1$, then C faces a choice between 10 and $9 - 2\varrho_2$, which implies that a pure-strategy, subgame perfect, psychological Nash equilibrium has A trusting B and neither B nor C choosing to behave corruptly. So, we have verified the intuition in the *Huai Nan Tzu* that the supervisor's behavior can make a real difference. A second psychological equilibrium reproduces the nonpsychological equilibrium of A never trusting, $p_1 = \varrho_1 = \pi_1 = 0$, and $p_2 = \varrho_2 = 0$. There is also a strictly mixed-strategy equilibrium with $0 < p_1 = \varrho_1 < 1$ if $11 - 2\varrho_2 - 2\pi_1 \geq 10$, or $\varrho_2 + \pi_1 \leq \frac{1}{2}$. Let us try $\varrho_2 = \pi_1 = \frac{1}{4}$. In the above game, if $p_1 = \pi_1 = \frac{1}{4}$, C faces a choice between 10 and $11 - 2\varrho_2 - \frac{1}{2}$. This means that $p_2 = \varrho_2 = \frac{1}{4}$ is a possible solution. If $p_2 = \varrho_2 = \frac{1}{4}$, then B faces a choice between $(0.25)(10) + (0.75)(-2) = 1$ and $2 - 4\varrho_1$. For B to be indifferent between those choices, $p_1 = \varrho_1 = \frac{1}{4}$ is required. But, in this case if A trusts with any positive probability, A's payoff is $-\frac{3}{4}$ multiplied by the probability that A trusts. Thus, in conclusion, there are three psychological equilibria: A trusts, $p_1 = \varrho_1 = 1$, and $p_2 = \varrho_2 = 1$ with payoffs of (10, 10, 10); A sometimes trusts, $p_1 = \varrho_1 = \pi_1 = \frac{1}{4}$, and $p_2 = \varrho_2 = \frac{1}{4}$ with payoffs of (0, 0, 0); and finally, A never trusts, $p_1 = \varrho_1 = \pi_1 = 0$, and $p_2 = \varrho_2 = 0$ with payoffs of (0, 0, 0). It is straightforward to construct a general version of this game (as we did in Figure 3) and verify that there will be three such equilibria for a reasonable set of parameter values. As long as the voters (A) trust government leaders (B), those leaders can make a genuine difference by being sincere and upright ($p_1 = 1$), because then government officials (C) will serve the people honestly ($p_2 = 1$). This verifies the maxim from the *Huai Nan Tzu*.

Organizational or corporate culture usually is referred to as the set of shared beliefs about how the organization will react to circumstances as they arise (see, for example, Milgrom and Roberts, 1992; Kreps, 1990). In our framework of analysis, organizational culture is captured by the expectations π_1, π_2, ϱ_1, and ϱ_2 which are verified to be correct ($\pi_1 = \varrho_1 = p_1$ and $\pi_2 = \varrho_2 = p_2$) in equilibrium. The organizational culture of no corruption ($\varrho_1 = p_1 = 1$ and $\pi_2 = \varrho_2 = p_2 = 1$) can be supported by the presence of expectation-dependent emotions, as we have demonstrated above. Everybody benefits materially from this kind of an organizational culture. No single player is able to achieve material gains from a unilateral defection to being unremorseful. Note the difference between these results and those in the last section. In Section 2, even if A chooses to trust, there are still three possible psychological equilibria, of which two involve corrupt behavior. With the three-tier hierarchy discussed in this section, we can treat Bs as the leaders of a government or the managers of a company and Cs as the government officials or the rank-and-file workers. We can then proceed to analyze the internal workings of the government organization or private corporation as a two-tier hierarchy.

In such an organization or corporation, leaders or managers can have an important impact on the organizational culture through creation of shared expectations by their exemplary behavior. We show that when leaders commit to being virtuous and trustworthy by choosing $p_1 = 1$, the unique psychological equilibrium involves all officials and all workers choosing to serve the public honestly. But, if leaders choose to behave corruptly by choosing $p_1 < 1$, then so do their subordinates.

The social norms discussed in Section 2 and the organizational cultures discussed in this section provide means for societies and organizations to maintain order without more law. We have demonstrated the effect of norms or cultures on the development of cooperative behavior. Such norms or cultures are sustained by the existence of expectation-dependent remorse, which serves the strategic role of making everybody better off materially. In other words, this expectation-dependent emotion allows for the implementation of a Pareto improvement, not just in terms of emotional payoffs but also in terms of observable material payoffs.

4. Related Literature and More Examples

We have provided models of how social order can be maintained without further law. Our model focuses on the role of emotions as first- and second-party sanctions. In a related article (Huang and Wu, 1992), we model the impact of such emotional responses as pride, anger, and disappointment on decisions about suit, settlement, and trial in litigation. Frank (1988, 1993) and Hirshleifer (1987) explain how emotions can ensure the credibility of threats or promises, while Hirshleifer (1993) models the logic of the affections and passions in economics. The rational-actor model in modern economics is usually devoid of any emotions as motivations of human behavior. Regret, guilt, remorse, shame, and embarrassment are usually seen to be signs of irrationality or, at best, nonrationality. If they are even recognized, emotions are viewed as sustaining behavior motivated not by rationality but by compliance to social norms. Elster (1989) even distinguishes between outcome-oriented and nonconsequentialist motivations, and proceeds to classify social norms as proscribing non-outcome-oriented behavior. Thus, he seems to view individual human behavior as motivated by two distinct types of motivations, differentiating between rationality and social norms. Instead of viewing social norms as being in conflict with rational behavior, we propose that social norms be understood as sustained by emotions that make compliance with those norms rational. We have shown how expectation-dependent remorse can overcome the problem of trust in societies and organizations.

Tax compliance is a phenomenon that naturally fits into a principal–agent framework, as in Reinganum and Wilde (1985) or Graetz, Reinganum, and Wilde (1986). The literature on evasion also provides a good example of our analysis because of the finding that people are more willing to pay their "fair share" of taxes if they believe that others are not cheating the government. Baldry (1986) reports on two experiments that indicate tax evasion is not viewed as just amoral gambling to maximize expected returns, but instead is

influenced by moral considerations. Pyle (1989) offers further evidence that decisions regarding tax evasion are not made independently of each other and incorporates the notion of social stigma into formal theoretical models. Gordon (1989) provides models that explain why some taxpayers never evade. Cowell (1990) reviews empirical support and constructs theoretical analyses suggesting that inequity, inequality, injustice, morality, disgrace, and conformity to convention all matter in tax evasion. In fact, he notes that, "the concept of 'stigma' . . . presumably involves a fairly complex interaction between a person's evaluation of the consequences of his own decisions and how he thinks such decisions would be viewed by others in the light of what they are doing—in other words, his perception of other people's perception of his own actions" (p. 108). He goes on to state that "a person's propensity to dodge taxes seems to be strongly affected by the number of other people who are already doing the same" (p. 108). More broadly, Grasmick and Green (1980) empirically test for what roles fear of informal sanctions by peers and moral commitment in the form of internalization of legal norms, in addition to formal sanctions by the law, may play as deterrents to criminal behavior. Schelling (1978) discusses numerous instances of a related idea, that of critical mass, examining behaviors ranging from turning on your headlights at dusk to crossing against the light. The difference is that in his examples, people behave in a certain way because there is safety in the number of others who act similarly, while in our analysis, people act in similar ways because they experience certain emotions that depend on their expectations about the number of others acting similarly; it so happens that the definition of equilibrium expectations requires that they be self-fulfilling. The two types of motivations can coexist, as they do in Schelling's story of the "dying seminar," where there is no point in attending a seminar nobody else attends, but there also is no shame or embarrassment from not going to a seminar that nobody thinks (correctly) that anybody else will attend.

5. Conclusions

This article has formulated psychological game-theoretic models of how to mitigate problems of opportunistic behavior in principal–agent and three-tiered hierarchical relationships. An interdisciplinary approach to studying how trust or its lack makes or breaks cooperative relationships is provided by Gambetta (1988). Our analysis provides formal support for the intuition expressed in the *Huai Nan Tzu* and several of the contributions in Gambetta (1988) concerning the important roles that social norms and organizational cultures play in controlling corruption and maintaining order without more law.

The law and economics literature teaches us that enforcement of laws is quite costly, in terms of the resources employed by the police, judges, courts, jury trials, and prisons. In addition, there is the age-old question raised by a Latin poet, "Who will take care of the caretakers?" The looting and rioting in south central Los Angeles that followed the Rodney King trial verdict is only the latest example and reminder that external control mechanisms are often

quite ineffective once people believe that others are disobeying laws and/or violating social norms. Ultimately, any order without law that is achieved by the development of social norms has to be maintained by organizational cultures that enjoin people not to abuse trust, conditional on trust being honored by others.

Human beings possess a very strong emotional desire not to be suckered. It is not good for a person's self-image to be taken advantage of by fellow humans. All of us would like to view ourselves in as positive a light as possible. Sometimes this is accomplished by self-delusion; however, if we want to assume rationality of expectations in our models, optimization under such deception should be ruled out. We observe in our models that the remorse arising from betraying trust might just as easily be replaced by the pride derived from honoring trust. In terms of the formal models, it makes no difference whether betraying trust causes remorse and bad feelings or honoring trust leads to joy and good feelings. Our conclusions are not changed by such a change of perspective. However, due to well-known framing effects and related cognitive limitations in the mental accounting of humans, the difference between avoiding negative feelings and seeking positive feelings might be significant.[9]

In the future, we plan to investigate generalizations of these models that incorporate asymmetric information and repeated play. These are important extensions because all of the models in this article involve only one-shot play and symmetric, complete information. The assumption of common knowledge about payoffs is standard in the psychological game theory literature because a theory of asymmetric information psychological games has not yet been developed. But ever since the pioneering attempts by Kreps, et al. (1982) to explain seemingly irrational business practices (such as collusion in a finitely repeated cartel) in terms of asymmetric information games, it has become standard to assume some type of asymmetric information in industrial organization models. We believe that incorporating the possibility that people can fake a conscience or pretend to behave morally is important, because it would require the introduction of "types" of players with different sensitivities of emotions to expectations. In our models, any member of a population of individuals possesses emotions that depend on their expectations to exactly the same degree as all others in that population. This is obviously a strong assumption. It simplifies the analysis greatly and assumes away such interesting behavior as deception or signaling of emotional responses. The assumption of a single-shot game similarly rules out interactions between emotions and observed behavior that can result in a theory of loyalty developed over time in response to unexpectedly cooperative treatment. Both asymmetric information and repeated play can explain a great deal. Indeed, as we mentioned in Section 2, the folk theorems demonstrate the ability of a little bit of the right kind of uncertainty or repetition to explain any individually rational, feasible behavior. Huang and Malueg (1992) prove folk theorems for psycho-

---

9. See Tversky and Kahneman (1981) for more on the framing effect.

logical games that demonstrate the ability of certain expectation-dependent emotions to explain any feasible behavior.[10] It might seem there is no point in adding time or asymmetric information to psychological game-theoretic models because of the folk theorems. But, such a viewpoint misses two important points: first, organizations persist over time and involve decentralized information; and second, answering (or even asking) the questions of how social norms and organizational cultures are formed in the first place and how they evolve requires more than a single-play, complete-information setting.[11]

Although we did not study the origin of emotions or norms, we do not endorse the anthropological functionalist explanation that they exist because they are good for the group. In our models, the moralistic or emotional behavior in question is not just psychologically rewarding to individuals but is materially so, because the material gain to behaving opportunistically would not even arise were the individual known to be unremorseful. Of course, this argument rests critically on the assumption that payoffs—in particular, expectation-dependent emotional ones—are common knowledge.

Our article sheds light on the age-old question of how laws create order in a society. An answer based on deterrence theories presumes that potential lawbreakers know the precise fines and punishments associated with various crimes as well as the probabilities of enforcement. Our answer does not require such a heroic assumption because the route by which laws create and maintain order is through the creation or alteration of social norms. It also suggests that laws influence preferences by their effect on expectations about what is socially acceptable behavior. Thus, in conjunction with Cooter's (1994) thesis that decentralized lawmaking involves raising to legal status those community norms that are fair and efficient, our thesis is that decentralized order is accomplished by internalizing as social norms those laws that are just and perceived to be fair. In a related study (Huang and Wu, 1994), we consider how social norms about equality or equity are supported by expectation-dependent anger in bargaining over the design of institutions. In such two-stage psychological games, negotiations among prodemocratic forces occur in the first stage over bargaining power in the possible family of games to be played in the later stage. The approach in that study and in this one expands the domain of the rational-actor model by explicitly studying expectation-dependent emotions and, in so doing, overcomes Ellickson's (1987, 1989b) criticism that law and economics involves too limited a notion of rationality. Finally, Huang (1994) discusses the implications of expectation-dependent emotions for the notion of self and modern welfare economics in general.

---

10. Individual rationality of behavior is preserved because the modeler is allowed to change the payoffs to ensure this.

11. This might require the study of evolutionary or population equilibria in psychological games.

References

Ames, Roger T. 1983. *The Art of Rulership: A Study in Ancient Chinese Political Thought.* Honolulu: University of Hawaii Press.

Baird, Douglass G., Robert H. Gertner, and Randall C. Picker. 1994, forthcoming. *Game Theory and the Law.* Cambridge, MASS: Harvard University Press.

Baldry, J. C. 1986. "Tax Evasion Is Not A Gamble: A Report on Two Experiments," 22 *Economics Letters* 333–335.

Becker, Gary, and Stigler, George J. 1974. "Law Enforcement, Malfeasance, and Compensation of Enforcers," 3 *Journal of Legal Studies* 1–19.

Cooter, Robert D. 1994. "Structural Adjudication and the New Law Merchant: Decentralized Law for A Complex Economy." Presented at the fourth annual American Law and Economics Association meeting, Stanford University Law School, May 13.

Coursey, Don L., Elizabeth Hoffman, and Matthew Spitzer. 1987. "Fear and Loathing in the Coase Theorem: Experimental Evidence Involving Physical Discomfort," 16 *Journal of Legal Studies,* 217–248.

Cowell, Frank A. 1990. *Cheating the Government: The Economics of Evasion.* Cambridge, Mass.: MIT Press.

Craswell, Richard. 1993. "On the Uses of 'Trust': Comment on Williamson, 'Calculativeness, Trust, and Economic Organization'," 36 *Journal of Law & Economics* 487–502.

Dasgupta, Partha. 1988. "Trust as a Commodity," in D. Gambetta, ed., *Trust: Making and Breaking Cooperative Relations.* Cambridge, Mass.: Basil Blackwell.

Ellickson, Robert C. 1986. "Of Coase and Cattle: Dispute Resolution Among Neighbors in Shasta County," 38 *Stanford Law Review* 623–687.

———. 1987. "A Critique of Economic and Sociological Theories of Control," 16 *Journal of Legal Studies* 67–99.

———. 1989a. "A Hypothesis of Wealth-Maximizing Norms: Evidence From the Whaling Industry," 5 *Journal of Law, Economics, & Organization* 83–97.

———. 1989b. "Bringing Culture and Human Frailty to Rational Actors: A Critique of Classical Law and Economics," 65 *Chicago-Kent Law Review* 23–55.

———. 1991. *Order Without Law: How Neighbors Settle Disputes.* Cambridge, Mass.: Harvard University Press.

Elster, Jon. 1989. "Social Norms and Economic Theory," 3 *Journal of Economic Perspectives* 99–117.

Frank, Robert H. 1988. *Passions within Reason: The Strategic Role of the Emotions.* New York: Norton and Company.

———. 1993. "The Strategic Role of the Emotions: Reconciling Over- and Undersocialized Accounts of Behavior," 5 *Rationality and Society* 160–184.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory.* Cambridge, Mass.: MIT Press.

Gambetta, Diego. 1988. *Trust: Making and Breaking Cooperative Relations.* Cambridge, Mass.: Basil Blackwell.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological Games and Sequential Rationality," 1 *Journal of Games and Economic Behavior* 60–79.

Gordon, James P. F. 1989. "Individual Morality and Reputation Costs as Deterrents to Tax Evasion," 33 *European Economic Review* 797–805.

Graetz, Michael J., Jennifer F. Reinganum, and Louis L. Wilde. 1986. "The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement," 2 *Journal of Law, Economics, & Organization* 1–32.

Grasmick, Harold G., and Donald E. Green. 1980. "Legal Punishment, Social Disapproval and Internalization as Inhibitions of Illegal Behavior," 71 *Journal of Criminal Law and Criminology* 325–335.

Greif, Avner. 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualistic Societies." forthcoming in *Journal of Political Economy.*

Hirshleifer, Jack. 1987. "On the Emotions as Guarantors of Threats and Promises," in J. Dupre,

ed., *The Latest on the Best: Essays in Evolution and Optimality.* Cambridge, Mass.: MIT Press.

————. 1993. "The Affections and the Passions: Their Economic Logic," 5 *Rationality and Society* 185–202.

Huang, Peter H. 1994. "Integrating Thinking and Feeling in Game-Theoretic Actors." Presented in the Microeconomic Theory & Mathematical Economics Seminar, Stanford University, May 12.

————, and David A. Malueg. 1992. "Psychological Remorse in the Prisoner's Dilemma and Folk Theorems for Psychological Games." A. B. Freeman School of Business Working Paper 93-ECAN-01, December.

————, and Ho-Mou Wu. 1992. "Emotional Responses in Litigation," 12 *International Review of Law and Economics* 31–44.

————, and ————. 1994. "The Role of Passions and Norms in Bargaining Over Constitutions." Presented at the Murphy Institute of Political Economy seminar on Constitutions and Constitutionalism, February 19.

Hwang, Kwang-kuo. 1987. "Face and Favor: The Chinese Power Game," 92 *American Journal of Sociology* 944–974.

Klitgaard, Robert. 1988. *Controlling Corruption.* Berkeley and Los Angeles: University of California Press.

————. 1991. "Gifts and Bribes," in R. Zeckhauser, ed., *Strategy and Choice.* Cambridge, Mass.: MIT Press.

Kreps, David M. 1990. "Corporate Culture and Economic Theory," in J. Alt and K. Shepsle, eds., *Perspectives on Positive Political Economy.* New York: Cambridge University Press.

————, Paul R. Milgrom, D. John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma," 27 *Journal of Economic Theory* 245–252.

Laffont, Jean-Jacques. 1990. "Analysis of Hidden Gaming in a Three-Tier Hierarchy," 6 *Journal of Law, Economics, & Organization* 301–24.

Macaulay, Stewart. 1963. "Non-Contractual Relations in Business: A Preliminary Study," 28 *American Sociological Review* 55–67.

Milgrom, Paul R. and D. John Roberts. 1992. *Economics, Organization, and Management.* Englewood Cliffs, N.J.: Prentice-Hall.

Pyle, David J. 1989. *Tax Evasion and the Black Economy.* New York: Macmillan.

Reinganum, Jennifer F., and Louis L. Wilde. 1985. "The Economics of Income Taxation: Compliance in a Principal-Agent Framework," 26 *Journal of Public Economics* 1–18.

Rose-Ackerman, Susan. 1978. *Corruption: A Study in Political Economy.* New York: Academic Press.

Schelling, Thomas C. 1978. *Micromotives and Macrophenomena.* New York: W. W. Norton.

Tirole, Jean. 1986. "Hierarchies and Bureaucracies: Collusion in Organizations," 2 *Journal of Law, Economics, & Organization* 181–214.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice," 211 *Science* 453–458.

Williamson, Oliver E. 1993. "Calculativeness, Trust, and Economic Organization," 36 *Journal of Law & Economics* 453–86.