Building a Pathway to Cooperation: Negotiation and Social Exchange between Principal and Agent

William P. Bottom
James Holloway
Gary J. Miller
Alexandra Mislin
Washington University in St. Louis
Andrew Whitford
University of Georgia

This paper examines contracting between a principal and an agent from the perspective of both social exchange theory and rational choice theory. Two experiments were conducted that tested competing predictions from the two theories. The first study examined effort decisions made by an agent under a series of contracts that varied in social context and compensation structure. The second experiment examined the negotiation of a compensation scheme between a principal and an agent and the agent's subsequent contract fulfillment, to test the mediating effects of verbal communication between parties on contracting and contract fulfillment. Both studies yielded results consistent with social exchange theory. Exchange theory appears to provide a better basis for deriving principles of organization design than rational choice.

An inherent tension between the interests of an employer and an employee leaves their relationship vulnerable to conflict. Rational choice theorists have formally modeled this problem, seeking to characterize the most efficient form of compensation contract to manage the tension (Salanié, 2002). From a rational choice perspective, the inherent information asymmetry between principal and agent necessitates a contract in which the principal shifts risk to the agent to motivate an efficient level of effort. A principal will not generally be able to determine whether a given performance outcome (either good or bad) resulted from the agent's effort or from circumstances beyond the agent's control.

As Spence observed (1973: 370), “it is a general maxim in economics that people with the same preferences and opportunity sets will make similar decisions and end up in similar situations.” Spence's maxim also holds for agency theory. Because the principal is not generally able to determine how the agent's effort contributed to the performance outcome, the contractual solution predicted by agency theory must be based on what is observable—the outcome itself. By crafting a compensation scheme that makes the agent's compensation heavily dependent on the outcome, the principal can be assured that the agent's self-interest motivates a high level of effort. A principal will not generally be able to determine whether a given performance outcome (either good or bad) resulted from the agent's effort or from circumstances beyond the agent's control.

As Spence observed (1973: 370), “it is a general maxim in economics that people with the same preferences and opportunity sets will make similar decisions and end up in similar situations.” Spence's maxim also holds for agency theory. Because the principal is not generally able to determine how the agent's effort contributed to the performance outcome, the contractual solution predicted by agency theory must be based on what is observable—the outcome itself. By crafting a compensation scheme that makes the agent's compensation heavily dependent on the outcome, the principal can be assured that the agent's self-interest motivates a high level of effort, even without monitoring the agent's actions. Because employees are generally far less diversified financially and less able to bear risk than employers, however, the use of this type of incentive contract represents an inherent sacrifice in efficiency. A rational principal determined to elicit hidden effort from rational agents must pay a premium to the employee to make him or her willing to take on risk that the principal is better able to bear (Harris and Raviv, 1979). This is a one-size-fits-all solution that does not depend on actors’ estimations of each other’s motivation or trustworthiness. If a principal could discern the agent's true intentions, this information would certainly be useful in managing relations with the agent. But discerning intentions is simply not possible. An agent's promise of high effort in exchange for a beneficial, low-risk contract is simply cheap talk; because agents can easily renege on such promises, they will not be credible to the principal and so cannot affect the outcome (Crawford and Sobel, 1982).

© 2006 by Johnson Graduate School, Cornell University.
0001-8392/06/5101-0029/$3.00.

We gratefully acknowledge the financial support of the Russell Sage Foundation and the Weidenbaum Center on the Economy, Government, and Public Policy. We thank Dan Brass, Stuart Bunderson, Leda Cosmides, Kurt Dirks, Claus Langfred, Chris Long, Judi McLean Parks, Ryan Quinn, Ray Sparrowe, Elke Weber, and three anonymous reviewers for constructive comments on earlier versions of this manuscript. We also greatly benefited from the expert editorial advice of Linda Johanson.
The basic predictions of agency theory have received little empirical support. In a survey of compensation practices, Baker, Jensen, and Murphy (1988) observed that pay systems actually contain far less variable pay than agency theory predicts. Studies of sales-force compensation plans (Eisenhardt, 1989; Coughlan and Narasimhan, 1992) have found substantial fixed pay features that are inconsistent with agency theory. Aggarwal and Samwick (1999) argued that executive pay may at least be trending toward greater reliance on variable pay and becoming more consistent with agency theory, but a recent meta-analysis of studies of the pay of chief executive officers (Tosi et al., 2000) found no empirical evidence to support such a pattern. Most of the variance in compensation was related to firm size; firm performance accounted for less than 5 percent of the variance, though particular firm variables moderated these relationships. Attempts to align top executive pay are particularly weak when ownership is dispersed, creating a “managerially controlled firm” (Tosi and Gomez-Mejia, 1989; Hambrick and Finkelstein, 1995).

Field tests of agency theory, however, have been inherently limited by the absence of experimental control and randomization (Tosi, Katz, and Gomez-Mejia, 1997). Experiments that have been conducted on the negotiation of compensation contracts have also found marked deviations from agency theory predictions. In experiments coupling the joint establishment of a wage and an outcome-contingent bonus (Conlon and McLean Parks, 1990; McLean Parks and Conlon, 1995; Miller and Whitford, 2002), principals have generally provided agents with high fixed wages and bonuses that were, theoretically, insufficient to motivate a high effort from a rational self-interested agent. Yet the agents in these experiments have responded by providing very high levels of costly effort. The result has been efficiency in exchange that exceeded the maximum achievable through agency theory solutions. Coupled with the field studies, these experiments indicate that the basic model of rational, self-interested choice underlying agency theory may be seriously flawed or incomplete. Design prescriptions based on this overly simplified theory may be misguided at best, counterproductive at worst. Social exchange theory (Homans, 1958; Thibaut and Kelley, 1959; Emerson, 1976) provides a potential explanation for the empirical failures of agency theory.

Emerson (1976: 336) defined social exchange theory as a “frame of reference” regarding “actions that are contingent on rewarding reactions from others.” He pointed out that both operant learning and rational choice had been used to provide alternate “starting points” for a theory of social exchange. From this perspective, agency theory represents a unique form of social exchange theory with particular assumptions about preferences, judgment, motivation, and the nature of the bargaining process. In his review of the literature, Emerson did not address Trivers’ (1971) alternative model of the psychological underpinnings of social exchange, one based on evolutionary reasoning. Though exchange theorists are still only beginning to explore the implications of this alternative conception of social exchange (e.g., Kameda,
Takezawa, and Hastie, 2003; Kelley et al., 2003), a growing body of evidence is consistent with Trivers’ (1971) notion that humans have evolved unique capabilities to reason about social exchange. These capabilities lead people to reason differently about social decisions than they would about problems that do not involve interdependence. Making social exchange salient in contract negotiations is likely to result in outcomes not predicted by agency theory.

This paper examines contracting between a principal and an agent from the perspective of both this conception of social exchange theory and rational choice theory. Two experiments were conducted that tested competing predictions from the two theories. The first study examined effort decisions made by an agent under a series of contracts that varied in social context and compensation structure. The second experiment examined the negotiation of a compensation scheme between a principal and an agent and subsequent contract fulfillment by the agent. This design provided an opportunity to test the mediating effects of verbal communication between parties on contracting and contract fulfillment.

SOCIAL EXCHANGE IN CONTRACTING

Social exchange theory encapsulates the special case of market exchange studied by economists, exchange that is “carried out under special circumstances and with a most useful built-in numerical measure of value” (Homans, 1958: 598). The theory has developed through the cumulative contributions of anthropologists (Boas, 1910; Malinowski, 1922; Mauss, 1954; Otterbein and Otterbein, 1965; Lee, 1979; Boehm, 1984; Kent, 1993), sociologists (Blau, 1955; Gouldner, 1960; Emerson, 1976), and social psychologists (Thibaut and Kelley, 1959; Jones, 1964; Adams, 1965; Kelley et al., 2003) who adopted the view that social interaction “is an exchange of goods, material goods but also non-material ones, such as the symbols of approval or prestige. Persons that give much to others try to get much from them, and persons that get much from others are under pressure to give much to them” (Homans, 1958: 606).

Agency theory can be seen as a form of exchange theory based on particular assumptions associated with rational choice. Homans (1961, 1983) sought to derive a general theory of social behavior from a purely behaviorist model of operant learning. In his view, exchange theory merely elaborated the social implications of operant conditioning. But the limitations of operant models of complex learning are now strikingly evident. The ease with which certain kinds of associations can be learned is clearly species specific (Revusky and Garcia, 1970). Humans have developed particularly acute capabilities for processing information and learning certain kinds of associations regarding social interactions (Saxe, Carey, and Kanwisher, 2004).

Learning theorists have shifted away from the simple associationism that defined Homans’ psychological theory toward a more modular conception of mental functioning. Fodor (1983: 37) characterized modular cognitive systems as “domain specific, innately specified, hardwired, autonomous, and not assembled.” Trivers (1971) described how evolutionary pres-
sures shaped the development of one such “complex, regulating system”—the psychological system underpinning what he called “reciprocal altruism” in social exchange. This form of altruism is the behavioral tendency to adhere to the norm of reciprocity in varied contexts and over varying time frames.

The psychological system that underpins exchange has a number of facets, including acute abilities to detect cheating, the emotional system of liking and disliking that supports the formation of friendships that evolve out of a shared history of reciprocal altruism, and moralistic aggression as a response to perceived cheating. The term cheating encompasses the outright failure to reciprocate benefits received and more subtle attempts to shade the terms of exchange in one’s own favor, reciprocating less than the benefit that was originally received. Judgments of trustworthiness are associated with the problem of detecting those likely to behave opportunistically in an exchange.

Specific emotional states, such as anger or rage, are subsystems closely related to the detection of cheating or the perception of unfairness in the distribution of outcomes from exchange partners. The resultant emotions have been linked to physiological reactions such as insomnia (Greenberg, 2006) and to spiteful behavior that damages both parties’ material interests (Bies and Tripp, 1996; Bottom, Eavey, and Miller, 1996; Pillutla and Murnighan, 1996; Sanfey et al., 2003). By overriding cognitive systems that enable choice based on careful cost-benefit calculations, anger and arousal facilitate moralistic aggression, the punishment of cheaters that would not arise from a system based purely on rational calculation.

Cosmides and her colleagues (Cosmides and Tooby, 1992; Tooby and Cosmides, 1996; Stone et al., 2002) provided striking experimental evidence of the existence of the acute abilities to detect cheating that Trivers had earlier hypothesized. They found that most subjects they tested demonstrate considerable aptitude for solving logic problems when the description of the problem is embedded in the context of detecting violations of the rules of social exchange. Similar subjects struggle when the same problem is described in purely abstract terms or even when it is embedded in a different type of meaningful context. Other studies have shown, as Trivers predicted they would, that the strong emotional reactions that follow the perceived detection of cheating triggers costly reprisal or moralistic aggression (Bottom, Eavey, and Miller, 1996; Pillutla and Murnighan, 1996; Price, Cosmides, and Tooby, 2002; Sanfey et al., 2003).

A social exchange theory based on this evolutionary model departs from the predictions of rational choice in a number of ways that have implications for the agency problem and organization design. One of the most basic of these differences is with respect to the motivating properties of a fixed wage. According to agency theory, a payment that is made irrespective of effort exerted or outcomes produced cannot affect the agent’s effort. According to exchange theory, however, agents who perceive a fixed wage as a benefit guaranteed by
the principal will feel an obligation to repay the debt that accompanies the receipt of a good or service and will provide a high effort to discharge the obligation. A cycle of obligations incurred and repaid will over time deepen trust and affection between the principal and the agent. This should facilitate future mutually beneficial exchanges.

Although preferences are complete and fixed in rational choice theory, they are tied to interpersonal history in exchange theory. Affection, goodwill, or friendship develops endogenously from a history of beneficial exchange (Homans, 1961; Trivers, 1971; Bottom, Eavey, and Miller, 1996; Tooby and Cosmides, 1996; McCabe and Smith, 2001; Kelley et al., 2003). History shapes the perception of rewards and consequences in subsequent exchanges as well as trust between the parties. Violated expectations of exchange that have negative consequences for a party create ill will and punitive sentiment, shifting preferences toward retribution and evening the score (Boehm, 1984; Bies and Tripp, 1996; Price, Cosmides, and Tooby, 2002; Labianca and Brass, 2006).

The two theories also differ with respect to the function and impact of verbal communication. In exchange theory, the sources of reward that may flow from an exchange are potentially quite broad. In an employment relationship, these include monetary compensation, intrinsic rewards that flow from doing the work, fringe benefits, and various symbols of status (Homans, 1961; Adams, 1965). They can also encompass “affection, love, formal courtesies, expressions of fairness, fair value (as in merchandise), and reliability (as part of the purchase of a service). Insult, rudeness, and rejection are the other side of the coin” (Adams, 1965: 278). Jones (1964) found a number of regularities in the way that various communication tactics (other-enhancement, opinion conformity, and self-presentation) either provided a source of value or influenced perceptions of the costs and benefits incurred in an exchange.

According to exchange theory, various forms of talk play important roles in motivating effort. The effect may be direct, through the provision of the kind of value that generates reciprocity. It can be indirect, shaping perceptions of the motivation and trustworthiness of the other party. In rational actor theories, communications that correspond to a show of affection, formal courtesies, or expressions of fairness, fair value, reliability, insult, rudeness, or rejection are considered “cheap talk” that should have no impact on the strategic behavior of rational actors. In theory, a rational agent interested in receiving the most pay for the least effort could readily claim otherwise in a verbal exchange without suffering any penalty. Recognizing that these statements lack credibility, the party’s actions will not be influenced by them. Yet experimental evidence is consistent with the notion that fixed wages can motivate high effort. Fehr and his associates have published a series of experiments (Fehr, Kirchsteiger, and Reidl, 1993; Fehr, Gächter, and Kirchsteiger, 1997; Fehr and Falk, 1999) indicating that an exchange theory based on motives other than simple rational choice is necessary to explain contracting. Fehr, Gächter, and Kirchsteiger (1997)
simulated a labor market in which the employer makes a non-retractable wage offer, after which the employee makes a one-time labor effort decision. In some treatments, the employer and employee were exogenously paired, though in other treatments some employees were able to select which employer’s wage offer to accept. These institutions created an artificial form of incomplete contract, given that the wage offered could not be made contingent on effort. No form of punishment was available to employers were the employee to actually provide a low effort level. The game-theoretic prediction in both cases is for the employee to provide the lowest effort level and for the employer to offer a wage that barely covers the cost of that effort. Yet employees frequently provided more than the minimal effort in both the exogenously paired and competitive market treatments. The level of effort that employees provided increased with the size of the wage premium that they were offered. Employers evidently expected such behavior, because many of them consistently provided a much higher than minimal wage even in one-shot, anonymous transactions.

Rational decision makers are highly sensitive to the revenues, costs, and likelihoods associated with different courses of action. They are sensitive to the opportunity cost associated with a given investment. Rational decision makers should demonstrate invariance to the descriptions of the problems, focusing only on the underlying consequences of the actions (Arrow, 1982; Tversky and Kahneman, 1986). In the modular conception of social exchange theory, the mind comprises specialized mental systems for handling very specific tasks, including the task of managing the complex accounting problems associated with ongoing social interaction. This theory predicts that individuals’ choices will not demonstrate invariance; they should reflect sensitivity to cues about the context that are provided by the description of the problem. Cues indicating that a risky choice is embedded in a social exchange will trigger different systems for reasoning and potentially quite different behavior. A favorable balance between gains and losses received through social exchange should trigger perceptions of obligation that motivate actions to discharge the obligation. Unfavorable balances constitute not simply an unwise investment, they represent an opportunity to detect a would-be cheater. The first experiment here examined the impact on investment behavior of cues about an exchange.

**EXPERIMENT 1: THE SOCIAL CONTEXT OF DECISION**

Experiment 1 tested whether subjects would show invariance to descriptions by manipulating the way in which decisions were described to participants. Participants were presented with a series of risky decisions similar to those faced by the agent in Miller and Whitford’s (2002) study. In the investment condition, the description of the problem contained no cues to a social context for choice and of exchange with another party. The description contained only the basic parameters of a risky decision. No reference was made to effort, to a contract, or to another party who could be affected by the choices. Each alternative was described as a risky prospect with different monetary payoffs and probabilities.
One component of each alternative was a fixed return—a monetary sure thing that could take on one of eight possible values (0, 1.75, 3.5, 5, 6.75, 8.5, 10.25, or 12). The second component was a variable return, described as a specified probability of a high return, which assumed one of six possible values (3, 6, 9, 12, 15, or 18), with a complementary probability of zero return. For each pair of gambles, participants chose one of three alternatives: to invest $5.00, to invest $8.50, or to forego investing altogether. A larger investment increased the probability of receiving the positive variable return. With a $5.00 investment, $p = .50$ was the chance of receiving the high variable return. The $8.50 investment increased the probability to .80.

From a rational choice perspective, the decision sets in the exchange condition were identical to those in the investment condition. We refer to this manipulation as the frame or context variable. The verbal descriptions of the problems were modified to provide cues that there was a social context for the risky decisions. In the exchange condition, these alternatives were described as a series of contract offers made by fictitious owners. That these were merely hypothetical owners, not real people, was stated several times in the instructions. We included additional descriptive cues in the instructions that this risky choice was part of a social exchange. We gave hypothetical owners distinctive names for each problem using unisex names to dampen any possible random noise introduced by gender effects. Names were randomly assigned to contracts, with different assignments for each participant, to insure that choice patterns were not biased by any inadvertent associations with particular names. The fixed return was described as a “flat wage.” The variable return was described as a performance “bonus,” contingent on a successful but uncertain project outcome. The set of choice alternatives was described as “action low” (at a cost of $5), “action high” (at a cost of $8.50), or “action quit,” rejecting the offer to work on the project (at a cost of $0). Figure 1 illustrates how the several cues in the exchange condition provided a different description of the choice problem than the investment condition by comparing the two choice sets available on the computer screen displays of one particular risky decision. Participants were asked to make a choice and confirm their decision.

The instructions in the exchange condition emphasized that the owners listed did not actually exist and were used only as a means of describing choice scenarios. These hypothetical owners stood to gain more from a successful project outcome than an unsuccessful project outcome. As in the investment condition, the likelihood of success was 80 percent with a high effort that cost the agent $8.50. The likelihood of success was only 50 percent with a low effort that would cost the agent $5.00. Project success was 0 percent if the agent elected to quit.

Agency theory predicts that choice under risk depends on probability and monetary payoffs that are invariant across problem descriptions. From this perspective, a costly large investment depends entirely on the expectation of the variable return. A variable return of at least $11.67 was neces-
Figure 1. Sample decision problems as they were presented to participants in Experiment 1.*

Investment choice condition

<table>
<thead>
<tr>
<th>INVESTMENT</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Return</td>
<td>$12.00</td>
</tr>
<tr>
<td>Variable Return</td>
<td>$6.00</td>
</tr>
</tbody>
</table>

Chance of Variable & Fixed Return

<table>
<thead>
<tr>
<th>Your Choices</th>
<th>Chance of Fixed Return ONLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invest $8.50</td>
<td>80%</td>
</tr>
<tr>
<td>Invest $5.00</td>
<td>50%</td>
</tr>
<tr>
<td>No Investment</td>
<td>0%</td>
</tr>
</tbody>
</table>

Social exchange condition

<table>
<thead>
<tr>
<th>CONTRACT</th>
<th>2</th>
<th>FROM SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat Wage</td>
<td>$12.00</td>
<td>Bonus</td>
</tr>
</tbody>
</table>

Cost of Action | Chance of $30 Project | Chance of $10 Project

<table>
<thead>
<tr>
<th>Your Choices</th>
<th>Cost of Action</th>
<th>Chance of $30 Project</th>
<th>Chance of $10 Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action HI</td>
<td>$8.50</td>
<td>80%</td>
<td>20%</td>
</tr>
<tr>
<td>Action LO</td>
<td>$5.00</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Action Quit</td>
<td>$0.00</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

* Participants clicked on their choices on the computer screen and then were asked to confirm their decisions.

necessary to compensate a rational individual for the extra cost of a high effort. An extra 30 percent chance at $11.67 just provided an expected gain equal to $3.50, the marginal cost for making a high rather than low investment. This variable return yields an expected value of $(11.67)(.80) – ($8.50) = $0.84 in the event of high investment and $(11.67)(.50) – ($5.00) = $0.84 in the event of low investment. Participants are not predicted to make high investments (or effort) in risky
choices that offer variable returns (or bonus) of less than $11.67, regardless of the fixed return (or flat wage). Rational choice also predicts that participants will choose no investment (or quit) whenever the expected value of an investment is negative.

Social exchange theory predicts that the pattern of choices will vary predictably as a function of the problem context. The system of reasoning employed to make choices will differ in the two conditions. Because the description of the problem provides no cues to an exchange context in the investment condition, cognitive processing should proceed without triggering the complex psychological system that regulates social exchange. In this condition, social exchange theory would predict that choices will approximate rational choice predictions. Individuals will choose a high investment if and only if the variable return exceeds $11.67. The size of the fixed return should have no effect on the willingness to make a high investment. When the expected value of the investment is negative, participants will forego the investment altogether.

For participants in the exchange condition, the verbal cues regarding contracts, counterparts, and effort should trigger the operation of the psychological system of reasoning that regulates social exchange. If the fixed return is described as a wage offered by a hypothetical owner, then the hypothetical owner has provided a benefit to the recipient. By accepting the offer, the participant will have incurred an obligation to reciprocate by providing a benefit to the hypothetical owner. In the experiment, the only way for the agent to reciprocate is by providing a high rather than a low investment in effort. Social exchange theory predicts an interaction between the problem context and the fixed return on an agent's decisions. In the investment condition, no cues are provided that would trigger exchange reasoning, nor would the fixed return be translated into a debt that must be repaid.

**Participants and Procedures**

We conducted all experimental sessions in the experimental laboratory of a business school at a private Midwestern university. A total of 33 undergraduate students participated, including 20 males and 13 females. The students ranged in age from 18 to 22, with a median age of 19. We recruited these participants through a campus-wide advertisement billed as an opportunity to earn money in return for a one-hour session involving business decision making.

Upon arrival, we randomly assigned subjects to a computer terminal and to either the investment condition or the exchange condition. Sessions varied in size from 9 to 13 participants. After reading the instructions, participants completed a quiz testing their comprehension. In the exchange condition, one of these questions insured that all participants were aware that the owner mentioned in the problems was fictitious. We allowed subjects to begin the task only after they had successfully answered all of the questions. Here they were asked to make a series of decisions about 48 separate contracts presented in random sequence at their computer terminal.
The instructions informed participants that at the end of the experiment, one of the 48 contracts would be randomly selected. Their pay for the experiment would be based on the decision they made with respect to that contract. If a contract was randomly selected and the subject had chosen to quit, then the participant would earn nothing for the task. If the contract called only for a flat wage of $10, then his or her earnings would be $10 less the cost of the level of investment he or she chose to make. If the contract included a nonzero variable return, then the participant’s compensation would depend on the project’s success. Project success was itself determined at the end of the experiment by random number generation. The odds of success or failure were made contingent on the agent’s decision to make either a high or low investment.

Once all subjects completed these decision exercises, but before the outcome was realized, they were offered an additional $10 to complete an additional questionnaire. Included in this questionnaire were items assessing demographics and risk tolerance. Each participant agreed to complete the questionnaire. One effect of this offer was to scale up all subjects’ payments, insuring that even those subjects who invested $8.50 in an unsuccessful project were able to earn a positive return from the experimental session. Participants were not informed of this scaling until after they had accepted the risks and made their decisions.

Results

Problem context or frame had a significant impact on the agent’s decisions about effort. The upper panel of table 1 shows the investment decision frequencies pooled over subjects and contracts as a function of problem context. Decisions to quit or to make a high investment were more common in the exchange context. Sixty-four percent of all decisions to make a low investment were made by participants in the investment condition. Those working under the exchange condition were less likely to provide what was characterized as a low amount of effort toward project success.

The lower panel of table 1 breaks down the tabulation of subjects’ actual choices by the choice that would maximize the expected monetary payoff. Whether quitting, providing low effort, or providing high effort, the choice that maximized expected monetary payoff was the modal decision of participants in both the exchange and investment conditions. When the payoffs favored a low effort, participants in the exchange condition were actually more likely to choose one of the options that did not maximize the expected payoff, either by quitting (.19) or by providing high effort (.36). Subjects in the investment condition made many more low investment decisions than did those in the exchange condition.

To model investment levels as a function of condition and payoffs, participants’ choices were coded as 0 = no investment, 1 = low investment, or 2 = high investment. Zavoina and McKelvey (1975) developed a model for the analysis of such ordered categorical response variables (see also Greene, 1997). We estimated two ordered logistic models.
using Huber-White standard errors adjusted for clustering to correct for the correlation of residuals within subjects. The first model predicted the participant’s investment level first as a function of fixed return and variable return. The second model added frame or context as a third predictor, coded as 0 = investment condition and 1 = exchange condition. This model also incorporated the interaction terms between fixed return and context and between variable return and context. Table 2 summarizes estimates for these two models. Variable return and fixed return both predict participants’ resource investment. Frame and the fixed return by frame interactions are also significant predictors when added to the model.

Table 1

<table>
<thead>
<tr>
<th>Problem Context</th>
<th>Decision</th>
<th>Investment</th>
<th>Exchange</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled</td>
<td>Quit</td>
<td>0.44</td>
<td>0.56</td>
<td>335</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0.64</td>
<td>0.36</td>
<td>533</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.46</td>
<td>0.54</td>
<td>716</td>
</tr>
<tr>
<td>N</td>
<td>816</td>
<td>768</td>
<td>1584</td>
<td></td>
</tr>
<tr>
<td>When Quit maximizes expected payoff*</td>
<td>Quit</td>
<td>0.89</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0.11</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.00</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>85</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>When Low maximizes expected payoff†</td>
<td>Quit</td>
<td>0.12</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0.64</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.24</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>323</td>
<td>304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>When High maximizes expected payoff‡</td>
<td>Quit</td>
<td>0.08</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0.31</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.61</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>408</td>
<td>384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* When the expected payoff of choosing high or low effort is negative, given the contract terms, the wage is zero and the bonus is $9 or less, or the wage is $1.75 and the bonus is $3 or $6.
† When the expected payoff of low effort is positive and greater than the expected payoff for high effort, given the contract terms, the wage is $3.50 or greater, or the wage is $1.75 or greater and the bonus is $9 or greater.
‡ When the expected payoff of choosing high effort is positive and greater than the expected payoff of choosing low effort, given the contract terms, the bonus is $12 or higher given any wage.

Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable return</td>
<td>0.251***</td>
<td>0.233***</td>
</tr>
<tr>
<td>Fixed return</td>
<td>0.371***</td>
<td>0.299***</td>
</tr>
<tr>
<td>Frame</td>
<td>–1.310*</td>
<td></td>
</tr>
<tr>
<td>Fixed return × Frame</td>
<td>0.180**</td>
<td></td>
</tr>
<tr>
<td>Variable return × Frame</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>MU (1)</td>
<td>2.760</td>
<td>2.252</td>
</tr>
<tr>
<td>MU (2)</td>
<td>5.220</td>
<td>4.739</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>–1199.470</td>
<td>–1182.300</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.280</td>
<td>0.292</td>
</tr>
</tbody>
</table>

* p < .05; ** p < .005; *** p < .0001.
Discussion
The pattern of results violates the rational choice principle of descriptive invariance. A rational choice “depends on the opportunity set from which the choice is to be made, independently of how that set is described” (Arrow, 1982: 6). Yet in this experiment, the description of the choice problem as a problem of investment or a problem of exchange had a significant effect on participants’ behavior. Rational choice predicts the variable return should determine investment decisions independent of the fixed return or problem context. The fixed wage could rationally influence the decision to quit or provide a low level of investment. Only a sufficient variable return could rationally motivate a high investment.

Choices in the pure investment context, without exchange cues, generally reflected this agency theory logic. Participants did show a somewhat greater than predicted tendency to choose a low investment even when the high investment maximized their expected monetary payoffs. This might reflect some tendency by more risk-averse investors to hedge. Exchange condition participants showed a very different tendency, consistent with the predictions from social exchange theory.

The fixed return had a significant effect on an agent’s investment when that return was referred to as a wage paid by a hypothetical other person. Not only was the interaction of context with wage significant, but the main effect for context was also significant and negative when entered in model 2. The sign reflected the greater tendency of those in the exchange context to quit rather than provide an expected value-maximizing low effort. Participants in the investment context had little difficulty making a low financial investment if the fixed return was high but the variable return was not promising. At least some of the participants in the exchange context appeared to treat the decision to take the contract as a kind of obligation to provide maximal effort, even at a financial cost. Rather than shirk to make it financially rewarding, many chose to quit or to provide high effort, even when shirking was the expected value-maximizing choice.

The process of reasoning about investment decisions varied considerably when cues indicated that the decision was part of a social exchange with another party. This pattern of context-dependent choice is consistent with the proposition from social exchange theory that a fixed wage constitutes a benefit provided by the principal to the agent that carries an obligation to reciprocate with an expenditure of high effort. That obligation has motivational force independent of the calculus of risk and return. The appropriate scale for interpreting whether a fixed wage constitutes a sufficient benefit to merit effort clearly depends on the context—on the costs incurred and the returns realized by the different parties to the exchange.

This experiment permitted a comparison of the processing of information in the two contexts. The results were consistent with exchange theory, though not with agency theory predictions. The study had a number of methodological limitations, particularly the absence of a genuine exchange between
interacting parties. We sought to address these limitations and more closely examine the cognitive operations underlying social exchange in Experiment 2. Rather than impose theoretically interesting contracts as in Experiment 1, Experiment 2 examined the process by which a principal and an agent discuss possible terms of employment. Prior research on bargaining and negotiation (see Bazerman et al., 2000, for a comprehensive review) has focused almost exclusively on understanding the forces that affect the fact and terms of agreement. The subsequent execution of the negotiated terms, the ultimate criteria for negotiating effectiveness, has been either automatically enforced by experimenters or simply assumed. By examining the agent’s effort decisions in executing the negotiated contract, Experiment 2 should also therefore contribute to understanding a vital but neglected dimension of bargaining.

By focusing on the development of a new relationship between two parties, rather than multiple decisions with hypothetical principals, the study permits a closer examination of the social cognitive mechanisms that shape perceptions of the other party, of the contract between them, and of the implementation of that contract. Specifically, the study was designed to test the impact that verbal communication between the two parties has on perceptions and on exchange behavior.

EXPERIMENT 2: CONTRACT BARGAINING

The experimental task context concerned the negotiation of an employment contract between a principal (owner) and an agent who would assist in undertaking a risky project. If the pair agreed on contract terms, the agent would then be called upon to make an unobservable decision about the effort that he or she would put into the project. A high effort decision increased the odds of the project’s success but cost the agent more financially than did a low effort decision. The contract negotiations required the determination of a fixed wage level and a bonus payable only in the event the project succeeded.

The owner could undertake the project only by hiring an agent to work on it. The agent could choose to provide either routine effort at a personal cost of $5.00 or high effort at a cost of $8.50. The probability of the project’s success was only 50 percent with routine effort but rose to 80 percent with high effort. The optimal contract for this problem is one that provides a performance contingent bonus of $11.67 in the event of the project’s success. With this optimal bonus and a zero wage, agency theory predicts that the agent should be willing to put out a high level of effort despite the extra cost.

The Size of the Pie

Agency theory and social exchange theory differ dramatically in other respects that we sought to test in Experiment 2. According to agency theory, the rational agent considers only the personal costs and benefits associated with alternative courses of action available. According to social exchange theory, agents attend to the equity of the exchange with the
principal. Varying the profits to the principal associated with a successful project, we sought to examine whether contracting behavior and the relationship between parties would vary as a function of the total, divisible value created by the exchange. This value is often referred to colloquially as “the pie.” The payoff that principals would earn in the event of a successful project (called the project payoff) was either $30 or $40. They would receive only $10 from a failed project. This treatment variable has no impact on the parameters relevant to calculating the principal’s optimal bonus: agency theory predicts a bonus of $11.67 regardless of the size of the pie. A rational agent will attend only to his or her possible payoffs and allocate personal resources accordingly.

From the social exchange theory perspective, the size of the potential pie should affect contracting. In elaborating Homan’s (1958, 1961) distributive justice proposition, Adams (1965) observed that individuals are acutely sensitive to the relative comparison of outcomes to inputs of the parties to an exchange relationship. They expect a balance between input and outcome. Agents who perceive an inequity may purposefully shift their own effort up or down to create a balance. Whether an agent’s input of effort is instrumental to the generation of $40 in profits or only $30 in profits will therefore influence the agent’s perception of equitable return for the effort invested. In the $40 pie-size condition, the payoffs provide sufficient return so that a principal could, in theory, provide the agent with a flat wage that covers the agent’s cost of a high effort, combined with a bonus large enough to provide a purely financial incentive for high effort. In the $30 pie-size treatment, a principal who attempted to do both would end up with a negative expected return.

Social exchange theory predicts that agents negotiating over the $40 pie condition will exert pressure for additional compensation to insure equity. This could take the form of either a higher wage, or a higher bonus, or both. If the principal provides this greater share, then it should lead to the provision of high effort. Should the principal behave as agency theory predicts, then compensation will be invariant with the size of the pie. According to social exchange theory, agents sensitive to distributive justice will respond by decreasing the input they control, i.e., the level of effort. Agency theory predicts no difference by treatment in either the terms of the contract or the agent’s effort.

Social Cognitive Factors

Though Miller and Whitford (2002) observed a great deal of behavior that was inconsistent with agency theory predictions, there was also considerable variability among subjects. They speculated that high wages reflected the principal’s trust in the agent, though they did not ask subjects about trust. In their study of wage setting, Kirchler, Fehr, and Evans (1996) examined effort choices in one-shot and repeated bilateral interactions, as well as a one-sided oral auction for labor when there was an excess supply of workers. They found that average wages and effort levels exceeded game theoretic predictions in each setting, but both varied around
that average: “some workers reciprocated high wage offers with high effort whereas others did not” (p. 327).

Kirchler, Fehr, and Evans attempted to explain some of this variability in behavior through questionnaire measures of motives and personality orientations such as introversion. Examining agents’ choices on the very last round in the repeated bilateral interaction, they interpreted responses to two questionnaire items (“Above all I tried to make a high profit,” and “I chose low effort levels in order to maximize my profit”) as reverse-scored indicators of an altruism motive. This reverse-scored measure of motives was negatively related to decisions to reduce effort on the very last round. A three-item measure of a “reciprocity motive” (“If I was offered a high wage I chose a high effort level,” “I always tried to be fair to my partner,” and “During the trading periods I was trying to be cooperative”) was moderately negatively correlated with decisions to reduce effort on the last round. A complex “reinforcement versus exchange orientation” construct was also negatively associated with the tendency to reduce effort on the last trial.

In Experiment 2 we sought to extend Kirchler, Fehr, and Evans’ (1996) approach to directly test agency theory predictions of contracting against exchange theory. The experimental design added the opportunity for participants to communicate verbally with each other prior to the exchange of offers. These messages were recorded, permitting an analysis of the content of the messages they exchanged. Our design also provided principals with a significant capability that they lacked in Kirchler, Fehr, and Evans’ (1996) study. They could construct contingent contracts in addition to or in lieu of offering a fixed wage to motivate effort. Adapting Kirchler, Fehr, and Evans’ attempts to explain variability in contracting behavior, we included several well-established and validated measures of affect and trust as part of the study.

The study also included a post-bargaining and agent decision questionnaire that assessed trust in the counterpart and the affect that participants experienced during the negotiation. Subjects completed it after the contract terms had been set by the owner and the agents had made their effort choices in secret, but before the outcomes were realized. The trust questions were adapted from the measure of trust in workplace relationships developed by Mayer and Davis (1999), who measured various facets of employee trust in management. Some facets, such as ability and outcome instrumentality are directly controlled by the experimental design. The agent’s ability, for example, is fixed by the marginal probability of the project’s success given a high as opposed to a low effort. Two facets that should be diagnostic for an exchange between a principal and an agent were included in the study.

Perceived benevolence, the extent to which a trustee is believed to want to do something good for the trustor, is a crucial mediating state in the social exchange theory logic we sought to test. The Oxford English Dictionary defines benevolence as “an expression of goodwill, an act of kindness; a gift or grant of money,” so it is the aspect of trust most closely tied to the principal-agent problem. If the flat wage
given by the principal is perceived as an act of benevolence, then it should stimulate the perception of an obligation to reciprocate. In the absence of that perception, there can be no obligation. We measured perceived benevolence by slightly adjusting the wording of the five-item benevolence scale from Mayer and Davis (“The principal is very concerned about my welfare”; “My needs and desires are very important to the principal”; “The principal would not knowingly do anything to hurt me”; “The principal really looks out for what is important to me”; and “The principal will go out of his/her way to help me”). In the interest of completeness, we also gave principals a modified version of these items that asked about their perceptions of the agent.

From the principal’s standpoint, the willingness to give value to the agent in the form of a fixed wage is itself an act predicated on a degree of trust. The principal must believe that the agent will appreciate what has been given, feel the obligation, and reciprocate with high effort. The four-item trust scale (“If I had my way, I wouldn’t let the agent have any influence over issues that are important to me”; “I would be willing to let the agent have complete control over my future payoffs”; “I really wish I had a good way to keep an eye on the agent”; and “I would be comfortable giving top management a task or problem which was critical to me, even if I could not monitor their actions”) was included to examine this hypothesis. We also gave agents a modified version of these items that asked about their trust in the principal.

Verbal Communication

Verbal communication between principal and agent, a very basic aspect of the employment relationship, has not been systematically examined in contracting research. Kirchler, Fehr, and Evans (1996) permitted participants to communicate only through offers and effort choices. Friendliness and other attributes of the parties could be inferred only from offers. Miller and Whitford (2002) allowed unstructured discussions between principal and agent but without recording or measuring any aspects of these conversations. They could not examine whether or how those communications influenced perceptions or contracting behavior. According to agency theory, such communications are irrelevant. Verbal claims made by either party are not strategically credible, so they should not influence the actions taken by rational self-interested counterparts. An agent who offers to provide high effort is making a claim that entails no cost and can be neither verified nor disproven by the principal. Cheap talk will be ignored by rational participants and so is predicted to have no effect on contract terms or agent effort.

But talk is not cheap, according to social exchange theory. Communication between parties shapes their perceptions of each other and their interpretation of subsequent actions. It provides an opportunity to begin low-cost exchange prior to contracting. Provision of non-material goods through conversation can take many low-cost, low-risk forms, including compliments, expressions of interest in the other side, and congenial banter. Social exchange theory would predict that principals and agents will use the opportunity to communi-
cate verbally to secure an employee. They will also use it to exchange conversation that begins the development of a relationship. Communication also provides the principal with an opportunity to ascertain the openness of the agent to reciprocity, to influence agents’ perceptions of the compensation package, and to build good will.

When offering a fixed wage, the principal may also attempt to insure that agents truly understand that they are getting much from the principal in the form of that wage and so therefore owe much to the principal in return. Failure to use available channels of communication to do this simply increases the chances of divergent interpretations of the contract and of inefficiencies in contracting. According to agency theory, the exchange in this experiment is a one-shot game without an opportunity for learning or relationship building. According to social exchange theory, an exchange of pleasantries, turn taking, and other non-monetary reciprocity can create a small history of exchange that serves as both practice and signal prior to the more risky one-shot contracting.

Because the negotiations were computer mediated, transcripts of the messages sent by each party were available for analysis. Three raters, who were blind to the experimental conditions and hypotheses, coded the communications in each dyad for various distinguishing features. Various attributes of the communications exchanged during the bargaining process were also coded by three raters. These attributes included “discussions of fairness” (0 = fairness of offers were not discussed; 1 = the fairness of offers were occasionally discussed; 2 = the fairness of offers were often discussed), “analytical calculations” (0 = no explicit discussions of expected returns or appeal to logic; 1 = appeal to logic but no explicit discussions of expected returns; 2 = explicit discussion of monetary expectations and expected return from contracts), “expressions of trust” (0 = no mention of trusting or trustworthiness; 1 = explicit verbal reference to trusting one’s counterpart); and “small talk” (0 = no discussions of matters unrelated to the task at hand; 1 = some discussion of personal matters not related to the task; 2 = much discussion of personal matters not related to the task).

Risk Preferences

The source of the efficiency loss predicted by agency theory is the assumption that principals are less risk averse than agents. To incorporate this key assumption into our study, we measured risk preferences and assigned subjects who were relatively more risk averse to agent roles and those who were less risk averse to principal roles. Risk preference was measured by presenting each subject with a series of 12 paired choices to estimate the subjects’ certainty equivalence for a 50-percent chance at $100. A similar method was employed by Murnighan, Roth, and Schoumaker (1988) and Bottom et al. (2000). Subjects who were relatively more risk averse were assigned to agent roles, and those who were less risk averse were assigned to principal roles.

Once subjects were assigned to roles, we also indirectly controlled risk preferences through the framing of the instructions. Instructions for the principals stressed that they were
negotiating over how much of the $30 or $40 project profit they would keep and how much they would have to give up. Instructions to agents stressed that they were negotiating over how much profit from a zero-profit reference point they would earn from their participation in the project. As predicted by the prospect theory value function (Tversky and Kahneman, 1986), prior research has confirmed that these asymmetries in negotiators’ framing lead to systematic differences in risk preference (Schuur, 1987; Bottom and Studt, 1993; Bottom, 1998). Loss framing contributes further to making principals relatively less risk averse than their gain-framed agent counterparts.

Agency and Social Exchange Theory Hypotheses

Agency theory predicts that subjects will not provide high effort in contracts that offer a bonus less than $11.67. Rational principals, in recognition of this fact, will create contingent contracts to induce high effort. The bonus will equal or just exceed $11.67 regardless of the project’s payoff. The flat wage, which redistributes income from the principal to the agent without any incentive properties, will equal zero. Verbal communication between the principal and agent prior to the exchange of contracts and the effort decision will be unrelated to decision making.

From the social exchange perspective, the agent’s response should be sensitive to more than the level of the bonus. If a fixed return is perceived as a benefit granted by a benevolent principal, then it will motivate a high level of effort. Pairs should achieve higher levels of efficiency in exchange, with the agent receiving a high fixed wage and reciprocating with a high level of effort. The agents’ felt obligation to reciprocate is an internal state mediating the relationship between the flat wage and the effort level. Discussion between the parties should also influence the felt obligation independent of the flat wage. This conversation involves an exchange of non-material goods that can also influence interpretations of other actions and motives, giving rise to different patterns of cooperation. In the principal-agent model setting, manifestations of friendliness through small talk should influence affect, the interpretation of the fixed wage, perceptions of the principal’s benevolence, and therefore the perceived obligation to reciprocate. The theory predicts that small talk and a gift of a large flat wage will affect perceptions of the principal’s benevolence. Perceptions of a benevolent principal will influence the agent’s effort.

The key response variable in this study was the agent’s choice of either high or low effort. In addition, we examined what predicts agent “voluntary cooperation”—when the principal gives a contract with a bonus of less than $11.67, meaning that the agent will maximize expected earnings by giving low effort, but the agent in fact gives high effort.

Methods and Procedures

A total of 112 undergraduate and graduate students, 67 male and 45 female, participated in the experiment. Their age ranged from 17 to 30, with an average of 21. Participants were recruited through an advertisement that billed the study

---

1 Because only three subjects chose “quit,” that option was not included in the analysis.
as an opportunity to earn money in return for participating in the one-hour session. All sessions were conducted in an experimental laboratory located in the business school of a private Midwestern university. Communication between participants took place through a local area network connecting personal computers in the laboratory. From eight to twelve people participated in any given session. The combination of session size and computer-mediated communication minimized the opportunity for subjects to learn the identity of their negotiating partner. As an additional check, we also inquired as to prior acquaintance with other participants in the session. Thirty-one percent of the subjects reported never having met any of the other participants who showed up for their particular session. Only 3 percent reported knowing two or more of the other participants in their session.

Upon their arrival, the experimenter read aloud to the subjects all of the instructions regarding the game and the method of payment. Participants then took a seat at one of the computer terminals and responded to a series of questions asking for demographic information and for choices between a set of hypothetical gambles and sure payments. The gamble was held constant, a coin toss for $100 or $0. The sure payments varied from $1 to $100 and were designed to assess the subject’s certainty equivalent over this range of payments at this point in time. We used scores on these items to get an index of risk aversion over monetary values in a range similar to those that would be at stake in the experiment. Other studies (Murnighan, Roth, and Schoumaker, 1988; Bottom et al., 2000) used similar assessments and found that they accurately predicted risk-taking behavior in a bargaining game. Once all the values were calculated, the program sorted participants into more risk-tolerant and less risk-tolerant groups. Those in the more risk-tolerant group were designated to be principals. The less risk tolerant were designated to be agents.

The negotiations began after the program randomly paired a principal and an agent. The program permitted open-ended communications between owner and agent until both mouse-clicked on a button indicating that they had a verbal agreement as to terms. The owner then was prompted to present a formal contract stipulating a wage amount and a bonus amount. The agent could accept this contract or reject it, terminating the experiment. Both parties understood that if the offer was rejected, then neither party would receive any earnings for the bargaining exercise. If accepted, then the agent was presented with a new screen asking him or her to decide between a more costly high effort or a less costly low effort. Subjects then completed a questionnaire about their perceptions of the bargaining process and their counterpart in the negotiation. The items included the PANAS measure of self-reported affect (Watson, Clark, and Tellegen, 1988) and Mayer and Davis’s (1999) benevolence scale.

Project outcomes were then realized based on the odds associated with the agent’s decision. Both agent and owner then learned the outcome of the project. The owner was informed only of the success of the project, not of the agent’s decision. Because of the structure of the payoffs, it
was conceivable that a participant could in fact sustain a monetary loss if the project failed. Therefore all participants were subsequently given an opportunity to participate in another questionnaire study on an unrelated subject in return for an additional $15 payment. Participants were not informed of this opportunity for additional payments until after the first experiment had ended.

Results

Two agents rejected the formal contract offered by the principal. Because they chose to quit, the four participants involved in these negotiations earned no money at all for this part of the experiment. The wage and bonus provisions of the full set of negotiated contracts varied considerably. Figure 2 plots contract terms and agents’ efforts with separate plots based on the size of the pie. Table 3 contains descriptive statistics on the key variables. The average bonus ($11.00) did not differ significantly from the agency theory predicted bonus of $11.67, but bonuses in the $40 pie-size condition ($11.89) exceeded those in the $30 pie-size condition ($9.06), t = 1.60, p < .05. The average fixed wage ($7.11) was significantly greater than the zero wage predicted by agency theory (p < .001), however, wages did not vary with the size of the pie. Two pairs negotiated a zero wage level. Both these contracts had incentive bonuses exceeding the $11.67 threshold. Both agents subsequently extended high effort.

Agents provided high effort in 40 of the 54 pairs. The bonus was less than $11.67 in 31 of them. In other words, 57 percent of the agents who chose high effort lacked a sufficient monetary incentive to justify that action, according to agency theory.

Predicting effort. We estimated a series of probit models to examine the factors predicting the agent’s decision to invest a high effort in the project. Because the response variable is highly skewed, we also estimated the models using complementary log-log regression. The results were similar, so only the probit models are reported. Table 4 summarizes the model estimates. In model 1, only the pie size and the contract terms are included as predictor variables. Bonus is a significant predictor of agent effort, but fixed wage is not. The higher the negotiated bonus, the more likely the agent was to provide high effort. Model 2 added the agent’s perception of the owner’s benevolence (mean = 13.9) as a predictor variable. Benevolence was a significant predictor; agents who reported that they believed the owner was more benevolent were more likely to contribute high effort.

Predicting non-rational or voluntary cooperation. Principals designed 43 percent of the contracts so that there were clear monetary incentives for the agent to choose high effort over low. In fact, 21 of the 23 agents working under a contract with a bonus in excess of $11.67 provided high effort. This provides strong support for the effect of bonuses on outcomes, an important aspect of the agency theory logic. Principals designed 57 percent of the contracts such that agents should have chosen a low effort level if they were seeking to maximize their own expected payoffs. Only 12 of
these 31 relatively risk-averse agents actually did choose low effort. Nineteen violated agency theory logic.

We examined an additional model to understand the origins of this voluntary or “non-rational” cooperation behavior among those not playing according to an agency theory strategy. The last column of table 4 shows the probit results predicting effort among only this subset of cases. Neither the pie size nor the bonus had a significant effect on the decision to give high effort within this restricted subgroup of contracts. The contractor’s perception of the principal’s benevolence did predict high effort. A contractor who believed that
the owner was interested in his or her welfare was more likely to invest in high effort without an adequate monetary incentive to do so.

Predicting benevolence. Using OLS estimation, we examined a model of the contractor’s perception of the owner’s benevolence. Table 5 displays the results. Model 4 includes only the contract terms as predictors. Both flat wage and bonus coefficients were significant. The higher the flat wage the owner was willing to pay, the more benevolent he or she appeared to be to the contractor. Model 5 adds an additional set of predictors, coded aspects of the conversation exchanged between the principal and the agent. These codes included the level of small talk engaged in by the owner (0 for none, 1 for some, and 2 for much small talk), the level of analytic calculation discussed by the parties (0 for none, 1 for some, and 2 for much), and the level of concern for fairness expressed by the owner (0 for none, 1 for some, and 2 for much). Three raters, blind to the hypotheses, read through the transcripts of the bargaining and coded them. Table 6 provides fragments of conversations illustrating each of the codes.

The addition of the owner-small-talk variable also improved predictions of the owner’s perceived benevolence. The more
small talk the owner engaged in, the higher the perceptions of benevolence. The other coded variables did not influence perceptions of benevolence. Model 6 included the self-reported negative affect of both the owner and the agent as additional predictor variables. This too improved predictions of the owner’s perceived benevolence. The more negative affect the owner experienced, the less the agent perceived him or her to be benevolent. The more negative affect the agent experienced, the less the agent perceived the owner to be benevolent. Negative affect is commonly the product of feelings of inequity or unfair treatment in social exchange (Adams, 1965; Witvliet, Ludwig, and Vander Laan, 2001).

Predicting the principal’s contract offer. Given the variation in contracting terms, we also examined models that would predict the principal’s decision to offer a fixed wage. The four-item trust scale from Mayer and Davis (1999) had lower internal consistency in this experiment than in their original field study (alpha = .58). Based on an item analysis, we added responses to an additional question, “Overall how trustworthy would you consider your counterpart in this experiment?” This augmented five-item scale had increased internal consistency (alpha = .68). It also explained 13 percent of the variance in the principal’s offer of a fixed wage, with \( b = .334 (.12), p < .01 \). The more trust the principal reported in the agent, the higher the fixed wage he or she offered.

Discussion
The results of Experiment 2 provided mixed support for agency theory. The agent’s effort was predicted by the magnitude of the outcome-contingent bonus. When provided with a sufficiently motivating incentive bonus, agents generally provided a high level of effort. Because of the relatively greater risk aversion of the principal in these dyads, they sacrificed efficiency to do so. Two pairs achieved the “second best” solution to the agency problem with a $0 fixed wage, a bonus above $11.67, and high effort expended by the agent. The results were inconsistent with agency theory in several
important respects. Most pairs negotiated a contract that was not of the agency theory form. These contracts had bonuses that were too small to align the interests of a rational, self-interested agent with those of the owner. Most had non-zero fixed wages. This represented a grant to the agent that theoretically should have provided no incentive to bear the additional costs associated with providing a high level of effort. Yet most of these agents violated that agency theory prediction by providing high effort anyway. In doing so, these pairs achieved better than the second-best solution to the agency problem.

It appears that there are two distinct pathways to motivating an agent to provide high effort. One approach, consistent
with agency theory, is to provide a sufficiently high level of incentive bonus. This aligns the monetary incentives of the agent with those of the principal and generally led to a high-effort decision in this experiment. This pathway constitutes a tradeoff between alignment and efficiency. It is a second-best solution because of the inherent inefficiency involved in shifting risk onto the party least prepared to bear it.

Most of the participants in this experiment took a second and quite different approach. This second pathway appeared to rely on the interpretation of the fixed wage as a “gift” from the principal to the agent, an expression of benevolence that obligated the agent to reciprocate. Effort varied with the size of the fixed wage. The fixed wage affected the agent’s perception of the principal’s benevolence. The principal’s decision to choose the first or the second pathway varied with trust in the agent. Principals who trusted the agent were more likely to choose the second pathway. They did not negotiate a bonus that provided a sufficient monetary incentive for agents to provide high effort. Yet by securing high agent effort without distorting risk sharing, they also attained a level of efficiency exceeding the second-best solution of agency theory.

Perceptions that the principal was benevolent were influenced by more than just the fixed wage. Aspects of the verbal communication between principal and agent, the total number of words exchanged, and the friendliness of those words (as rated by a neutral observer) were positively related to perceptions of benevolence. Two transcripts illustrate the way in which talk can influence subsequent exchange. The illustration of “small talk” in table 6 paired a wage of $5.00 with a bonus of $10.00 and verbal communication independently coded as friendly. The remarks included jokes and conversation about seemingly irrelevant issues like music concerts. The principal slipped in a reference to his or her personal need for earnings, a costly car repair, that framed his or her concessions to the agent as genuine sacrifices. The agent in this pair reciprocated the friendly tone of the conversation, rated the principal as very benevolent, and responded with high effort despite having financial incentives that were not aligned with the principal’s. A different pair negotiated an even higher fixed wage ($8.00), though their verbal exchange was not friendly:

**Principal:** Hello.

**Agent:** Hi . . . what flat wage are you willing to give me?

**Principal:** $8.50.

**Agent:** What is 30 times .80? 24?

**Principal:** Yeah . . . that’s right, what are you getting at?

**Agent:** If I choose action hi . . . then you will have an expected payment of $24 and my decision cost will be $8.50.

**Principal:** Correct, minus flat wage and bonus. I’m not going to offer you a large flat wage and watch you cut your costs with decision lo.

**Agent:** Let’s just figure out what the flat wage and bonus would have to be to get the same expected payoffs.

**Principal:** That’s what I’m figuring out now, or trying to at least.
Agent: If I do decision lo, then I will have a lesser chance of getting the bonus so it’s in my best interest to pick decision hi.

Principal: Right . . . so how about $8 flat and $8.50 bonus? That gives an expected payoff of $7.50 each.

This principal openly articulated the logic of agency theory: “I’m not going to offer you a large flat wage and watch you cut your costs with decision lo.” The professed lack of trust and the absence of any discussion of subjects other than the terms of the project evidently left the agent without a feeling of obligation to this principal. The agent subsequently provided low effort.

These two pairs show that a fixed wage does not provide a direct path to motivation. Reciprocity in social exchange is not a simple association of a principal’s stimulus triggering an agent’s response. The motivational force of a given wage depends on the agent’s perception of benevolence in the principal’s motives. According to the modular perspective, the mind is seen as a series of specialized systems, each expert at governing thought and action for particular kinds of problems. In the modular conception of social cognition, cooperation (in the form of a high effort in return for protection against risk) is not automatic, it must be constructed. Subjects brought considerable “mind-reading” skills to the task of inferring the intentions of their counterparts in these negotiations. Agents attempted to determine whether the principal was acting from a spirit of benevolence. The offer of a flat wage is just one indicator of that benevolence. While it may not be costly in financial terms, the exchange of verbal communication that preceded the offer also shaped perceptions. If agents did not perceive that a principal was benevolent, they were less likely to provide high effort.

GENERAL DISCUSSION

The two experiments indicate the merits of social exchange theory, a truly general framework for understanding agency problems and organizational design. Agency theory treats the exchange of effort for pay as a decision under risk. Yet people reason very differently about social exchange relations than they do about risky decisions with individual consequences. Experiment 1 demonstrated the context effects associated with cues that the decision was a social exchange. The mediating effects of benevolence in Experiment 2 provided additional evidence that the decision to provide effort is not simply tied to the self-image of someone who always gives a maximal effort. The decision to provide effort depends on perceptions of the other party in the exchange. Motivating high effort with a flat wage contract is neither infeasible nor an automatic form of reciprocity. It is not automatic because there is a very real human heterogeneity that subjects themselves are aware of: some partners are benevolent, some not; some partners are trustworthy, and some should not be trusted. Seen in this light, cooperation is not a foregone conclusion but a social accomplishment that depends on the give and take of signals and cues in the negotiation process.

Emerging evidence (Sanfey et al., 2003; Saxe, Carey, and Kanwisher, 2004) has begun to clarify the modular structure
of the mental systems involved in social cognition and the regulation of exchange. Agency theory fails to account for the distinctive architecture and operation of these systems. Most experimental participants in the second study acted as if they had an intuitive appreciation of the operation of these systems. The risk-sharing solution embodied in agency theory worked in this experimental context for the small minority of subjects who used it. But it worked less efficiently than did the relational approach based on trust and reciprocity.

The relational approach to contracting characterizes not just behavior in laboratory experiments but the basic form embodied in most business firms. The failure of most organizations to utilize pay practices prescribed by agency theorists (Baker, Jensen, and Murphy, 1988; Eisenhardt, 1989; Coughlan and Narasimhan, 1992) may reflect managers’ understanding that there is more than one way to motivate high effort. The pathway through benevolence and relationship building may be the more efficient pathway in many circumstances. Rather than attempting to achieve the second-best solution, which sacrifices efficiency for the alignment of monetary incentives, it appears likely that these firms are choosing to strive for higher levels of efficiency, for something closer to a first-best solution.

Aggarwal and Samwick (1999) suggested that executive compensation practices may be trending toward greater consistency with agency theory logic. If so, it is quite possible that at least some of these firms may be sacrificing the higher levels of efficiency possible from the second pathway through trust and goodwill. Rational choice theory helped to provide a more powerful theoretical foundation for what McGregor (1960) famously called the Theory X approach to management. Rather than treating employees as passive responders to stimulus-reward contingencies, it models them as being far-sighted and highly strategic utility maximizers of a purely self-interested orientation. The modular conception of social exchange theory provides an opportunity to create a more general theory incorporating both Theory X and the Theory Y practices associated with the human relations movement as special cases.

The ensemble of human resource practices commonly referred to as “high-commitment organization” (Miller, 1992; Pfeffer, 1994; Baron and Kreps, 1999) represents a coherent approach to utilizing the pathway from benevolence to motivation. This set, illustrated among others by Southwest Airlines, includes employment guarantees, egalitarianism in word and deed, self-managing teams, job enlargement, premium wages, extensive socialization and training, open sharing of information, and thorough screening of prospective employees. Many of these features (employment guarantees and flat wages) serve to shield employees from risk, rather than impose risk as part of a tradeoff to align incentives. Virtually all such features serve as ways to lead employees to perceive that the employer is benevolent. Any incentive compensation that is provided is based on team or unit performance, not individual performance.
This approach to design is certainly no panacea—it entails costs and certain risks (Baron and Kreps, 1999: 190). For employers, a prime risk of high-commitment organization lies in identifying and selecting employees. A failure to identify employees who will reciprocate their employer’s benevolence with high effort can prove fatal to the firm. At Southwest Airlines, much more time and energy is put into the selection process than at conventional firms (Pfeffer, 1994). And employees are always at risk of misreading the benevolence of their employers (Miller, 1992). In the selection process, selecting those more prone to reciprocal altruism becomes crucial.

Nor is high commitment management the only way to organize human resources effectively. Not every firm attempts to identify cooperative “types” as a precondition for employment, nor need they. Other firms are more concerned with identifying ambitious, risk-tolerant types who will respond to incentive-intense compensation plans. As in the experiments discussed in this paper, an employer that offers a sufficiently high contingent bonus can successfully motivate self-interested employees.

According to Spence (1973: 370), “it is a general maxim in economics that people with the same preferences and opportunity sets will make similar decisions and end up in similar situations.” Our claim of a contribution rests ultimately on a denial of this maxim. In experiments, as in employment relations, people are aware that those with the same preferences and opportunity sets may nevertheless be different—they may be more or less trustworthy, more or less compelled to reciprocate a benefit received with a comparable benefit given. Furthermore, they may differ in their mind-reading capabilities—their abilities to discern the intentions and motivations of others. We should expect heterogeneity in social exchange, because a subject’s willingness to cooperate is not unwavering but depends on that subject’s estimation of likely reciprocation in the other. There is no single way to organize for social action. Agency theory informs one mode of organization. For the majority of people, who are not only willing to reciprocate but are able to find partners in whom they detect the same willingness, much more is possible.
Negotiation and Social Exchange


