

# Reputation Formation in Simple Bargaining Games

Catherine C. Eckel

(National Science Foundation

and

Virginia Polytechnic Institute and State University)

and

Rick K. Wilson

(National Science Foundation

and

Rice University)

Paper presented at the Midwest Political Science Association Meeting, April 23-25, 1998, Chicago, Illinois. The National Science Foundation bears no responsibility for the conclusions drawn in this paper. The authors thank a whole bunch of people. You know who you are.

“Just put on a Happy Face!”  
-- Lee Adams and Charlie Strouse

## **Introduction.**

Legions of political advisors make their living suggesting how candidates should project themselves in public. This not only includes the symbols with which they are surrounded (the American flag, a spouse, children -- even a dog), but also includes physical appearance. Even leaders are subject to such advice. As recounted by Newt Gingrich, Trent Lott advised the Speaker to huddle closer to the microphone in a nationally televised speech so that the American public would not see how overweight he had become.<sup>1</sup> The advice offered to politicians echoes that given by our own mothers -- when meeting someone new, put your best foot forward to leave a favorable impression. This means being cleanly dressed, well groomed and wearing a smile.

Initial impressions are important for building reputation. While an initial impression does not form a complete reputation -- that is something built over time -- initial impressions are critical for determining how someone will be judged. The voluminous literature on stereotyping confirms this point.

If reputation is so important then it ought to be a central element of modeling human behavior. Some game theorists have been concerned with the role of reputation, especially when modeling repeated games (see for instance Kreps et al. 1982, and Fudenberg and Maskin, 1986). The work that has evolved out of repeated games focuses on the way in which actors, by observing the earlier actions of others, update their priors. However, an enormous hole in all of this literature has to do with the origin of priors.

In this paper we explore one way in which reputations are generated. We look explicitly at the ways in which actors form their initial priors about their counterparts. The settings we investigate are simple, one-shot, two-person sequential games. Each game has a unique subgame perfect equilibrium. At the same time the games contain an outcome that Pareto-dominates the equilibrium, but can only be reached by a sequence of moves that depends on actors trusting one another and engaging in reciprocal altruism. Both trust and

---

<sup>1</sup> For a sampling of interesting work on how simple physical cues affect the assessment of politicians, see Sullivan and Masters (1988), Masters and Sullivan (1989), Cherulnik et al. (1990), and Budesheim and De Paola (1994).

reciprocity are grounded in actors holding strong priors about their counterparts in the game; evolutionary psychology provides a basis for inferring the source of an actor's prior.

We first motivate this problem through a discussion of research in evolutionary psychology and its applicability to the equilibrium selection problem in game theory. Second we turn to a discussion of facial expressions as a source of initial reputation formation. We then present survey and experimental results that test the impact of facial expressions on expectations and behavior. The final section concludes with a discussion of avenues of future research and the importance of these findings for game theory.

### **Motivation**

Game theory tells us that rational actors will play dominant strategies, they can do no better than select strategies resulting in a Nash equilibrium, and subgame perfection is a powerful equilibrium prediction. Yet a series of empirical papers show that laboratory subjects eschew playing dominant strategies. These occur in ultimatum and dictator games (Camerer, 1998; Eckel and Grossman, 1996, 1998; Forsythe et al., 1994; Hoffman et al., 1994), investment trust games (Berg et al., 1995) and gift exchange experiments (Fehr et al., 1993). These results are not random or haphazard. Behavior inconsistent with game theoretic predictions is routine and patterned. This does not imply that game theoretic models are wrong: they often have considerable predictive power (for a discussion, see Ostrom, 1998). There are a variety of explanations for why game theoretic models fare poorly in some environments -- ranging from humans relying on simple heuristic decision rules, to learned patterns of behavior, to bounded rationality. One especially promising approach is suggested by Hoffman, McCabe and Smith (HMS, forthcoming) who argue that reciprocity explains the kinds of systematic anomalous results observed in the laboratory.

HMS first draw on results in game theory on reputation-based equilibria. That literature notes that many equilibria can be supported when players are engaged in infinite or finite-horizon repeated games and develop beliefs about the types of players in the game (see for instance Kreps et al., 1982 or Fudenberg and Maskin, 1986). In such games, it is simple for actors to update their beliefs about others by observing their play in a game. However, because there are many equilibria, problems of equilibrium selection arise. Drawing on evolutionary psychology, HMS propose that actors have evolved distinct mental modules for social exchange that serve as equilibrium selection mechanisms. Such mental modules are adaptive mechanisms that enable individuals to easily solve problems of multiple equilibrium or to settle on pareto-enhancing outcomes. These modules are adaptive in the sense that they are the product of millions of years of evolutionary

adaptation; the modules have not changed appreciably from the late Pleistocene when humans lived in small hunting and gathering tribes.<sup>2</sup>

In particular HMS settle on the concept of “reciprocal altruism” as a powerful mechanism for solving problems of social exchange. Evolutionary psychologists consider reciprocity to be a form of altruism, and altruism has long posed a serious problem for evolutionists. Self sacrifice looks like a quick path to extinction, yet seemingly altruistic behavior is pervasive in the animal world. Two theories have arisen. The theory of kin selection was proposed by Hamilton (1964) and elaborated by others: self sacrifice can be a successful strategy if, by sacrificing its own genes, the animal increases the reproductive success of its siblings, cousins, or other kin. This theory is supported by a great deal of evidence, but other apparently altruistic behavior does not support the theory. In some animal societies, self-sacrifice is observed among non-kin. This observation led Trivers (1971) to propose a theory of reciprocal altruism which takes the form: I give you food now with the expectation that you will give me food later. Such a discussion evokes economists’ notion of gains from trade.

Why might an evolutionary process select for reciprocal behavior? Reciprocity arises in animals who have the cognitive ability to keep track of who everyone is and what their behavior has been in the past and it does not require that animals consciously calculate the costs and benefits of cooperation. Reciprocity arises behaviorally from and is enforced by emotions. Gratitude, for example, is felt in proportion to the importance of the act, and acts like an IOU; guilt enforces repayment, and acts like a collection agency.

Reciprocal altruism can be a successful strategy in environments where there are gains to cooperation – “non-zero-sumness”, as Wright (1994) puts it. In these situations, reciprocal behavior in the form of a “tit-for-tat” strategy is evolutionarily stable against (small) invasions of free-riders (Axelrod, 1984). The tit-for-tat strategy requires cooperative initial play, but this can be a dangerously risky strategy when stakes are high. We might reasonably expect animals to develop mechanisms for judging when a potential partner is likely to be a good bet. How might such a mechanism arise for determining initial beliefs about a partner’s intentions?

Much of the focus on mechanisms leading to belief formation derives from evolved mental modules. On such module deals with “mind reading.” This is a shorthand way of referring to the process by which individuals draw inferences about the mental states of

---

<sup>2</sup> For a thorough discussion and standard statement see Cosmides and Tooby (1992) or Pinker (1997). Evolutionists study behavior in terms of its contribution to inclusive fitness – that is, the effect of the behavior on the organism’s success at getting its genes into the next generation. If an evolved mechanism for making decisions in a particular set of problems contributes more to inclusive fitness than other mechanisms, then it will out-reproduce alternative mechanisms, and will be “selected for”.

others. These inferences are ordinarily based on the words, actions and physical projections of others (see Baron-Cohen, 1995). Central to the ability to "read" the intention of others is the ability to quickly assess those who present a threat and those who do not. Le Doux (1996) summarizes research that shows that such perceptions occur very rapidly, more quickly than conscious, cognitive perceptions of a threat can be processed. HMS propose that such a module can determine the choice of initial strategy: trust and cooperate with "friends", and withhold trust and limit cooperation with those perceived as threatening.

HMS are careful to point out that this does not mean that mental modules are programmed ("hard wired") with a cooperative predisposition. All that is necessary is that humans are born with the capacity to learn such responses. Social exchange is an important component of social behavior, arguably sufficiently so to warrant the evolution of specialized attention mechanisms. There is now quite a lot of evidence that the human brain is predisposed to filter its environment to learn particular things, language perhaps most notable among them (Pinker, 1994).

### *Games.*

Two classes of games are examined in this paper. Both are two-person sequential games with perfect and complete information. These games are not designed to elucidate principles of evolutionary psychology. However, these games are interesting in that each has a unique subgame perfect equilibrium. At the same time they illustrate the potential gains from both trusting and reciprocal moves for the two actors.

To illustrate, consider the first game given in extensive form by Figure 1. Two actors, A and B, face a series of moves through the game and have the same preference ordering over outcomes:  $w > x > y > z$ . Actor A has the first move and can choose to end the game, with an outcome of  $y$  (and B also receiving  $y$ ), or he can pass the move to B. If passed, then B has a similar choice -- end the game with A receiving  $z$  and B receiving  $w$  or pass the move back to A who then chooses between equivalent outcomes.

<Figure 1 About Here>

The outcome to this game is straightforward. Under backward induction, B knows that she will obtain outcome  $x$  if she passes the move to A at the middle node (there is no strategic choice at the last node because all outcomes are equivalent for A). However, if B chooses to quit at the second node, then she will obtain her most preferred outcome,  $w$ . Her choice, then, will be to quit if the choice is passed to her. A, then, knows that he will

receive  $z$  if he passes the move to B at the first node. If A chooses to quit at the first node, then he will obtain outcome  $y$ . Because  $y$ , the third preferred outcome, is better than  $z$ , the least preferred outcome, A will choose to quit at the first node. That move constitutes the unique equilibrium for the game. What is worth noting is that both A and B are better off if A can trust B and B reciprocates that trust. If A passes to B and B in turn passes back to A, then both actors can gain their second-most preferred alternative.

The second game is similar and is given by Figure 2. Under backward induction at the last choice node A would choose down, gaining his most preferred outcome,  $w$ . At the second choice node B, knowing she will get outcome  $y$  if the move is passed to A, will choose to quit and gain her most preferred outcome,  $w$ . Therefore, at the first node A will choose to quit, obtaining his third-preferred outcome,  $y$ . Again A's best strategy is to chose to quit at the first node, and this constitutes the unique equilibrium for the game. In this case A gains his third-best alternative, while B obtains her least preferred outcome. Both actors would be better off if they could trust one another and if that trust was reciprocated.

<Figure 2 About Here>

While the first and second games predict the same choice by actor A, there are subtle differences between the games that are interesting for out-of-equilibrium play. For game one, if trusting and reciprocal behavior emerges, both actors do better than in equilibrium, and they both know what to expect. By contrast, in game two actor A not only has to trust B and believe that trust will be reciprocated, but B also has to trust A and believe that trust will be reciprocated at the final move. Although both actors are left better off at the final down move, yielding  $(w,y)$ , than at the equilibrium, B prefers the right move at the final node, which yields  $(x,x)$ . It is with this second game that out-of-equilibrium beliefs produce interesting implications for trusting behavior.

### **Reputation Formation**

Above we have noted the possibilities for gains to both actors *if there is trusting behavior*. But what is the locus of trust and how might it be communicated? In the discussion above we suggest that social signals can affect one's initial reputation and can be a possible source of trust. Our primary concern in this paper is with the ways in which initial reputations are formed and communicated.

There are many sources of possible cues for the kind of "mind reading" that is required to develop initial trust. Baron-Cohen (1995) emphasizes the importance of the eyes for signaling intentions and developing common understanding for individuals, this is

clearly not the only avenue for communication. Fridlund (1994) argues that human facial expressions provide critical cues for social behavior, though he is hardly the first to make the point. There is an extensive body of research on human faces and what they mean to observers. Much of this literature derives from Darwin (1872/ 1998), who argued that humans, like animals, have evolved patterns of signaling behavior, including (but not limited to) facial expressions.

The original thrust behind Darwin's characterization is twofold. First, he argues that facial expressions (and other display expressions) are evolved expressions that serve a function for the species. So a peacock's ostentatious display of his tail or a chimpanzee's baring of her teeth is something innate to the species. Second, expressions serve to signal something to others. In other words, universal, common signals are used to warn, invite or soothe members of the same species and on occasion other species as well.

Contemporary researchers largely follow the lead of Ekman (1972; 1983) and focus on the first part of Darwin's argument. With respect to humans this approach holds that there is a universal set of evolved human facial expressions. Many of these expressions are thought to be involuntary, and reflect basic emotions. The bulk of the research has turned toward understanding what facial expressions reveal about the underlying emotional state of the expressor. The literature has largely followed a research stream defined by Ekman (1972) in which he argues that universal expressions are the product of distinct emotions. As a consequence, facial expressions constitute emotional leakage, in which the emotional content should be obvious to others, since they too share the same universal repertoire of expressions. Of course, learned social behavior works to mask those emotions and cultural differences lead to different forms of masking. Therefore facial expressions can sometime be hard to read (for a general critique of the "universal" recognition of emotion, see Fridlund, 1994, Chapter 10).

Researchers for the most part have not taken up the question of the role of facial expressions in social signaling. Since the focus has been primarily on meaning and interpretation, there has been almost no investigation of behavioral responses. The consequences of facial expressions for behavior in different social settings is an open question.

This is a bit surprising, because there is an extensive literature on what facial expressions mean for children or for autistics. From the outset, faces appear to be crucial for child development. In an imaginative study Johnson et al. (1991) trace the reaction of new born infants to a paper stimulus about the size and shape of a human head. A variety of stimuli are used, ranging from an image resembling a human face to an image with the same parts, but scrambled, to a blank piece of paper. Measuring eye tracking and head

movement, these researchers find that newborns pay much closer attention to paper images resembling a human face than other images. These findings are all the more impressive in that the infants tested were less than one hour old. These researchers conclude that children are born with a system that orients them toward face-like patterns; only as they mature do they develop a cortical system that allows for sophisticated face-processing activities. As they note, "... a primary purpose of the first system is to ensure that during the first month or so of life appropriate input (i.e., faces) is provided to the rapidly developing cortical circuitry that will subsequently underlie face-processing in the adult." (p. 18).

Building the circuitry seems an important component of social signaling. In a study by Sorce et al. (1985), twelve-month old infants are shown to grasp quickly the difference between a mother's smiling face and an expression of alarm. The experiment used a plexiglass table, half of which was transparent, creating a "visual cliff" in the center. The infants were placed on the solid surface and at the clear end of the table stood the infant's mother. The majority of the infants crawled across the clear space when given a posed smile by the mother, whereas they did not if the mother produced a posed fearful expression. Obviously facial expressions evoke important social signals that are well understood at a very young age.

If facial expressions are an important part of social signaling, then what happens if an individual is incapable of reading another? Baron-Cohen (1995) contends that autistics lack the ability to read another's mind. He asks the reader to imagine how difficult life would be without the ability to read another's mind through facial expressions:

This is what it's like to sit round the dinner table.... Around me bags of skin are draped over chairs, and stuffed into pieces of cloth, they shift and protrude in unexpected ways.... Two dark spots near the top of them swivel restlessly back and forth. A hole beneath the spots fills with food and from it comes a stream of noises. Imagine that the noisy skin-bags suddenly moved toward you, and their noises grew loud, and you had no idea why, no way of explaining them or predicting what they would do next.(Gopnik, 1993, quoted in Baron-Cohen, p. 5)

In a series of experiments comparing normal with autistic children, Baron-Cohen and colleagues showed pictures of other children producing a number of emotional expressions (from happiness to sadness). While both groups of children could match basic emotions, autistic children made many more errors in matching pictures of surprised expressions, which Baron-Cohen argues are "belief-based" (p. 78). He notes that autistics appear to lack gaze-monitoring capabilities and so do not focus on the actions of others nor draw inferences about the intentions of others. This obviously interferes with their ability to conduct social exchange.

In a challenge to what he calls the "emotions view" of faces, Fridlund (1994) proposes a "behavioral ecology" view of faces, arguing that facial expressions and their



interpretation by others is crucial. "The balance of signaling and vigilance, countersignaling and countervigilance, produces a signaling 'ecology' that is analogous to the balance of resources and consumers, and predator and prey, that characterize all natural ecosystems." (Fridlund, 1994, p. 128). Facial expressions and their interpretation involve a delicate game in which expressions are signals about intentionality.

### *Abstract Images.*

Much of the literature using human facial expressions finds that particular expressions are difficult to "read." That is, the emotional content of an expression is often unclear (see the critique by Russell, 1993). Even something as simple as a "smile" can easily be misinterpreted or misrepresented -- especially if a single snapshot is pulled out of context (Ekman et al., 1998; Leonard et al., 1991; Fernandezdols and Ruizbelda, 1995)

Human facial expressions can be very complex; the muscle groups on the face can easily send a wide spectrum of signals. While researchers are looking for the six primary emotions -- happiness, sadness, anger, fear, surprise and disgust -- humans are capable of sending very subtle blends of expressions. In addition differences in physical attractiveness, slight differences in expression, and unfamiliarity with the posed face all lead to variations in assessing emotions. To correct for these problems a handful of researchers have adopted highly stylized aspects of faces in order to tease out the primary elements of facial expressions.

If there are specific components of expressions that signal specific emotional states, then these should be capable of being systematically evaluated. Taking this insight, McKelvie (1973) designed an experiments in which he used schematic representations of faces. These schematics resemble variations on the ubiquitous "happy face" wishing everyone a nice day. McKelvie used an oval to represent a head and then drew in line segment representations of eyebrows, eyes, nose and mouth. These were systematically varied and then presented as stimuli to subjects.

A total of 128 schematic faces were used; each subject was presented with a sample of 16 of those faces. Working one at a time, subjects were asked to rate how easy it was to find an adjective to describe the face and then asked to score the appropriateness or inappropriateness of each of 46 adjectives for describing the face. The adjectives reflected four different emotional categories (happy, sad, angry and scheming) and one other category (vacant). His analysis shows that the shape of eyes and the structure of the nose has little effect on evaluations. Instead, eyebrow and mouth shape have the greatest effect. He cautions that neutral (horizontal) eyebrow or mouth expression signals little.

"However, when brow and mouth move from the horizontal, clear differences in meaning

emerge: medially down-turned brows indicate anger or schemingness; medially upturned brows are seen as sad; an upturned mouth denotes happiness; and a down-turned mouth is seen as angry or sad.” (McKelvie, 1973, p. 345). In short, even simple schematic representations of faces can trigger emotional affect that is well recognized.

Part of McKelvie's study was replicated using pre-school children. MacDonald et al. (1996) used schematic drawings of facial expressions thought to represent the six primary emotions. At the same time selected photographs from Ekman and Friesen's "Pictures of Facial Affect" (1978) were used. In one of the experimental conditions children were asked to choose specific emotional categories when viewing either the pictures or the schematics. MacDonald et al. find that accuracy in picking the proper label was significantly greater for the schematic drawings than for the photographs. Accuracy varied, however, with children having the easiest time identifying happiness, sadness and anger (p. 383). The simplifications of the schematics were readily apparent to these children and the emotion evoked was usually readily interpretable. A similar finding for adults comes from Katsikitis (1997) whose subjects compare both pictures of actors and line drawings of those same faces. For certain of the emotions (like surprise), the line drawings tend to be easier to interpret (see also Yamada, 1993).

Aronoff et al. (1988) take schematic representation one step further. They create highly stylized stimulus displays that include objects like pairs of downward sloping lines and pairs of concave curves. These are highly abstract representations. Subjects are then given a series of items using 7 point semantic differential scales. The primary concern is whether some of the displays elicit a threatening subjective response. Indeed, Aronoff et al find strong effects associated with the different stimuli. As they note, the "primary visual configurations of angularity, diagonality, and curvilinearity examined in this study are quite different from the information that is customarily understood to convey the meaning of threat, for example, by eyebrows drawn together, by threatening gestures, or by angry words. These results are most interesting because they demonstrate that visual features that are, presumably, content-free, also possess the power to convey meaning to observers." (p. 654). This same point is driven home in a subsequent study by Aronoff et al. (1992) when they examine the degree of roundness and diagonality of simple geometric figures and the extent to which emotional affect is triggered.

The lesson to draw from these studies is that humans are very good at recognizing emotional content even in highly stylized schematics. Pictures have meaning and they are readily interpreted. Using these findings we move to two experiments to focus on the behavioral signals that are embedded in simple facial expressions. To avoid the ambiguity associated with still photographs of human faces we use stylized icons to represent facial

expressions. Based on the work of McKelvie (1973) and others, we limit our attention to the position of eyebrows and the shape of the mouth.

## Experiments

Two distinct experiments were conducted. The first is a survey designed to reveal subjects' impressions of a series of schematic faces. The survey measures the trustworthiness of the faces as well as the emotional affect attributed to them. The former is a departure from earlier work because it is designed to elicit how subjects' expectations about behavior are shaped by facial expressions. We use the results from the first experiment to select specific icons for the second experiment. The second experiment tests the effect of a subset of the schematic faces on the behavior of subjects in simple games involving trust and reciprocity. We are interested in knowing the extent to which the characteristics attributed to the faces are revealed in the behavior of subjects who have been assigned these same faces. In order to test this relationship, we select the faces from the first experiment that are judged to be the most distinct from each other.

*Experiment 1.* A survey instrument was administered to a sample of 524 subjects (324 male, 192 female and 8 who failed to indicate their sex) in Principles of Economics classes at Virginia Polytechnic Institute and State University in January, 1998. The classes consisted primarily of college sophomores; about 1/3 were business majors, 1/3 engineering majors and the rest from assorted fields. Subjects were asked to complete a three-page survey during a regular class meeting time, either at the beginning or the end of class, and were not compensated for their participation. On the first page of the survey, each subject was assigned one of the nine icons shown in Figure 3, and asked to rate its characteristics. (Only the data from the first page of the instrument are reported in this paper. That part of the instrument is reproduced as Appendix 1.)<sup>3</sup> The icons are based on a 3x3 design involving three manipulations of the mouth and three manipulations of the eyebrows, and were designed to evoke different affective responses.

<Figure 3 About Here>

---

<sup>3</sup> On the second page, subjects were asked to compare two of the icons on a series of characteristics. Ten face-pairs were distributed randomly among the subjects in each classroom. Instead of a semantic differential scale, subjects were given twenty-six descriptors and asked to check those matching each icon. The third page collected standard demographic data.

Subjects were randomly assigned to a particular icon and told that the icon “is supposed to represent a type of person.” They then were asked to choose the most appropriate response for their icon on twenty-five word-pair items using a seven-point semantic differential scale. In the scale, a value of (1) means the word on the left is “very” close to matching the meaning of the icon, (2) is “somewhat” close, (3) is “slightly” close and (4) is “neither.” The scale is symmetric to the right of (4). Left/right word order is randomly assigned for the word pairs, although for ease of exposition all words are represented so that positive traits are on the left.

In the analysis presented here we focus on ten paired items from the instrument: Five of the scale items relate to a behavioral assessment of the icon (does the icon reflect trustworthiness, generosity, cooperativeness, etc.) while the latter five items are common measures of emotional affect (does the icon reflect goodness, happiness, etc.). We have selected the four icons that are the most distinct from each other for further experimentation.

Figure 4 plots the mean response across the four selected icons for each word pair. Each icon’s means are connected with a line running down the graph. The icons at the top of the graph match the order of the means for the trusting/suspicious word pair. For example, the “happy” face is judged on average to be more trusting and trustworthy than the “sad” face, etc. What is clear from the figure is that the order is preserved across nearly all ten items. The icon on the left that we have labeled “happy” is always viewed as having more positive traits than the “devious” icon which is at the far right.

<Figure 4 about here>

The rough “eyeballing” of these means pretty much tells the story. Across the first five “behavioral” items, the differences across icons are significant under pair-wise t-tests (the only exception is between the “angry” and “devious” icons on the generous/selfish item). In other words, subjects perceive the icons as representing different behavioral traits. The “happy” icon is viewed as the most trustworthy, generous, cooperative and honest. The “sad” icon is viewed more positively than the “angry” icon, which in turn is viewed more positively than the “devious” icon across these behavioral traits.

The next five items constitute emotional affect items that have been used by a large number of researchers who study faces. Here a similar ordering is preserved, with the “happy” icon being the most positively evaluated. The “sad” icon is generally regarded positively, except on the happy/sad item, where it is appropriately rated by subjects. Finally, the “angry” and “devious” icons switch positions on the good/bad, kind/cruel and

friendly/unfriendly items. However, this is consistent with what we ordinarily think of angry and devious affective states.

Experiment 1 reveals that subjects differentiate between the various icons, and that the icons can be ordered from most to least “positive” on both behavioral and affective characteristics. These results are presented so as to set the foundation for experiment 2. There subjects are assigned to one of these icons. Their partner's icon type also is revealed. The results of experiment 1 allow us to predict the behavior of subjects in experiment 2. If subjects who are assigned different icons behave differently, or if the icon of a subjects’ partner affects his choice, then the perceptual results of experiment 1 are confirmed.

### *Experiment 2*

In the second experiment pairs of subjects participate in series of two-person sequential games. All games involve sequential choice under perfect and complete information. In a typical game each subject makes one or two moves. A total of 102 subjects were recruited from the local student population at Rice University. Students were contacted in their dining hall and asked to volunteer for a decision making experiment. Subjects signed up for one of fourteen planned experimental sessions.

The laboratory accommodates eight subjects, each seated in a cubicle formed by moveable partitions and facing a computer. Although subjects are in the same room and can hear one another, they cannot see one another’s computer screen. At the outset of the experiment subjects are cautioned not to speak and told that if they do so, then the experiment will be canceled. All experimental sessions are conducted over local area network and all communication between subjects is handled by the network.

Upon arriving at the laboratory subjects choose their seats and are asked to sit quietly until all the volunteers have arrived. At least nine subjects are recruited for each session in order to ensure that eight participate. If more show up, a volunteer is solicited and paid \$3.00 on the spot to leave. If no one volunteers, one subject is randomly selected, paid and dismissed. Once the requisite number of subjects appears, the experiment begins. In five of the 14 sessions fewer than eight turned up, and only six subjects participated in the experiment (only an even number of subjects could be used in this experiment).

Subjects are given self-paced instructions and shown how to indicate their choices in the sequential game.<sup>4</sup> These instructions are attached as Appendix 2.<sup>5</sup> In each

---

<sup>4</sup> The task for the subjects was referred to as a “decision problem.” The term “game” was avoided, but is used throughout the text to denote a decision problem characterized as a game theoretic problem.

experimental session, subjects participate in as many as 30 games.<sup>6</sup> Prior to each decision subjects are randomly rematched. Because of the limited number of subjects, same-pair play often occurs. However, subjects carry no unique identification in the course of the experiment, so it is impossible for subjects to know with whom they are paired in each game.

*Games.* A set of 18 distinct games is used during the course of the experiment. For the first six periods of play the games are presented in a fixed order. In all subsequent periods games are randomly selected for each pair in each period. Therefore in our analysis of this design we have to be careful to account for history effects that are a function of the order of play.

In this paper we analyze results from three of the games that have similar structure, as shown in Figure 5. Subjects observed exactly this game tree on their screen. Except that a player was identified as “YOU” at each decision node in which the player had a move.

<Figure 5 About Here>

*Procedure.* The same procedure is used in each period of play. Before each game begins, subjects are told which roles they are assigned (1 or 2) and who will move first. Each subject is randomly assigned to be player 1 or 2 in each period (in the experiment, we refer to “Decision Maker” 1 or 2). Information about a player’s counterpart is then revealed, depending on the icon manipulation as explained below. In the games we analyze here, player 1 always has the first move and player 2 waits until the first player’s choice is complete. At each of the first two nodes a player has two choices -- either to quit the game and take the payoffs or to pass to the next player. Once a choice is made, the other player is notified of the move. If the choice is to quit, the payoff box is circled on the computer screen, and both players are notified of the outcome and asked to record their payoffs for that period. If the move is passed, the second player must make a choice while the first player waits. The only communication between players is mediated by the computer; subjects are only told the moves of their own partners. If the game continues to the last decision node, the first player has the choice of two payoff boxes. Once the game ends,

---

<sup>5</sup> On average subjects took 7 minutes to complete the instructions, with no subject taking longer than 13 minutes. In post-experiment debriefings subjects remarked that the instructions were clear and said they had no problem with the task.

<sup>6</sup> In the first experimental session, because we were uncertain about the length of time required to complete the games, subjects participated in only 20 periods. In the second session, this was raised to 25 periods. In both sessions subjects were debriefed and asked whether the task was too boring or repetitious. Subjects indicated they were not bored and would not have minded participating in more decisions. All subsequent periods used 30 periods. Each game usually lasted less than a minute.

subjects are instructed to wait until all pairs have completed their own decisions. At that point the subjects are again shuffled and re-paired.

Participants were paid in cash and in private at the conclusion of the experiment. All payoffs in the games in Figure 5 are in U.S. dollars. Subjects were told at the outset that they would be paid only for a single period of play. At the conclusion of the experiment they were asked to draw one card from a deck of 100 electronic cards displayed on the computer screen. Subjects were told that each period had an equal probability of being chosen, and the algorithm for the selection ensured this. No matter which card was turned over, the program randomly selected a period and informed the subject of the period drawn and earnings. Subjects were asked to verify that payoffs for the period drawn matched what they had recorded. Before they were paid, subjects filled out an on-line questionnaire that asked them questions about their participation. No experimental session lasted more than 60 minutes and none was shorter than 47 minutes. On average subjects earned \$13.27 for their participation. One subject earned the maximum of \$30.00 and four subjects earned \$0.00 for their play. These latter subjects were paid a show-up fee of \$3.00, but not informed until their debriefing that they would be paid this amount.

*Icon manipulations.* Subjects know little about their partners in each trial. At best they know they are paired with other subjects in the same room, but they never know the identities of their partners. In order to build a reputational signal, however, each subject is assigned a permanent identity at the outset of the experiment. Five distinct manipulations of identity are used. In each session half of the subjects are randomly assigned one icon representation and the remaining half are assigned the other. Figure 6 presents the icon pairs used in the experiment. Subjects know there is a population of two types of icons in experiment. They know that each subject had been assigned an icon at the outset and that each player retains his or her assigned icon for the duration of the session.

We are very deliberate in not tying the icon to any personal characteristics of subjects. They are simply told the icon assigned to them will be theirs for the entire experimental session. Our sense is that this constitutes a very weak stimulus, and that a stronger connection between the icon and the subject would strengthen any observed behavioral effects.

<Figure 6 About Here>

The primary manipulations for the experiment involve the pairings of icons with human facial characteristics. As noted above and detailed below, we have explicit predictions about non-game-theoretic play in each of the games. These behaviors are

influenced by the icon type of a player and the partner. Four icons were used in which the angle of the eyebrows and orientation of the mouth were changed. The icon with downturned eyebrows and an upturned mouth (1) is characterized as “devious;” the icon with upturned eyebrows and an upturned mouth (5) is characterized as “happy;” the icon with downturned eyebrows and a downturned mouth (2) is “angry;” and the icon with upturned eyebrows and downturned mouth (6) is “sad.” The three pairs involved “happy and angry,” “devious and happy,” and “devious and sad.”

A second set of icons is used that has no human facial content. Here a rectangle (7) and an oval (8) are paired. These icons are chosen as one control condition for the experiment. It may be that subjects do not use the human facial content in the icons to select strategic play. Instead they simply view the world as consisting of two types – “them” and “us”. “In-group/out-group” effects are common in social psychological experiments (Tajfel and Turner, 1979; Turner, 1978), and we might reasonably predict that subjects identify with their own icon type and play differently then when confronting a different icon type.

In each trial a subject can be paired either with an individual with the same icon or an individual with a different icon. Prior to beginning a game subjects are shown the entire set of icons in the game (for an 8-person group this meant four images of each type of icon). When the subject is ready to begin, the icons appear to be shuffled on the screen and one is selected by the program. The screen then displays the subject’s own icon and the icon of his or her counterpart. When the subject chooses to continue, the game is then displayed.

A final control condition was also introduced in which subjects have no information about their counterparts. Their “icons,” for all intents and purposes, are blank (9). They are not presented any screens in which they are told about their counterpart’s identity, but simply play a series of games in which they are told they have been randomly matched with another participant. This control group is designed such that no reputational content is provided. Subjects in this condition should exhibit behavior that more closely approximates the predictions of standard models of game theory.

### *Predictions*

Now that the structure of the games and the experimental conditions are clear, we can state our hypotheses for this second experiment. The first hypothesis is taken directly from game theory. Despite the fact that subjects are assigned to different conditions -- some with and some without icons -- game theory would treat these as the same. The icons



(signals of initial reputation) are uninformative -- they amount to little more than "cheap talk."

*Hypothesis 1 (Game Theoretic):* Subjects will choose the subgame perfect equilibrium.

For games 1, 2 and 3 the prediction is that the first mover will choose to take the money at the first node and not pass the move to the second player.

A second hypothesis is derived from the earlier discussion of evolutionary psychology.

*Hypothesis 2 (Idle Evolutionary Psychology):* Subjects will engage in trust and reciprocal altruism.

This hypothesis draws on the idea that subjects are endowed with an evolved mental module that enables them to "read the intentions" of others. Being placed in the same non-threatening environment, subjects will work down through the sequential game. A trusting move by the first player will yield a reciprocity move by the second player. This hypothesis predicts no difference between groups assigned icons with faces, groups assigned icons that have no facial characteristics and groups assigned no icon. The strong prediction is that pairs of subjects will reach the third move, regardless of the condition to which they are assigned.

A third hypothesis also builds on evolutionary psychology. However, it also acknowledges the importance of facial expressions for triggering an initial reputation. This in turn provides sufficient information for reading the intentions of others.

*Hypothesis 3 (Reputation Formation):* With strong reputation priors about their partners, subjects will engage in trust and reciprocal altruism. Absent strong reputational priors, subjects will choose the subgame perfect equilibrium.

This hypothesis is more complicated than the second hypothesis. It picks up on the idea that trust is not always given. Instead, it is triggered by observations about a player's counterpart. We focus here on the reputational signal inherent in facial expressions -- even when those expressions are highly stylized. We know from the survey results in Experiment 1 that there are perceived differences across the different facial icons. These results allow us to predict that different facial icons will affect the degree of trust between two players with assigned icons. There are likely to be differences in trust and reciprocity moves even within the facial expression conditions. After all, if subjects are using these

expressions to establish initial reputations, then what the first mover sees in his partner and what he thinks the second mover sees in himself, will make a big difference.

The discussion surrounding hypothesis 3 leads to a number of explicit predictions:

- A. Subject pairs assigned to a condition in which icons have facial expressions will engage in trust and reciprocity moves at higher rates than conditions in which those expressions are absent.
- B. Subject pairs assigned to a condition with no icon or an icon without a facial expression should behave similarly -- the icons are uninformative.
- C. Counterparts assigned to "happy" or "sad" icons are more likely to elicit a trusting or reciprocity move.

The analysis that follows examines the different predictions derived from these hypotheses.

#### *Analysis.*

Tables 1-3 present summary data for each stage of the three games. The tables are laid out in a similar manner. The first column is a representation of the icon assigned to the subjects in the experiments. The column denoted "Own Face" contains decisions at that node of play for subjects who were assigned the icon in the first column. The column denoted "Other Face" contains the decisions at that node of play for subjects who were facing partners with that assigned icon. Table 1 contains decisions at the first node of the game; Table 2 contains decisions at the second node, which is reached only if the first mover "trusts"; and Table 3 contains decisions for the third and final node. For example, in Table 1, Game 1, 77.3 percent of subjects assigned the "devious" face chose to trust their counterpart at the first node of the game; subjects who faced "devious" counterparts trusted them 72.4 percent of the time.

It is apparent from the data in the first two tables that Hypothesis 1 can easily be rejected. About two-thirds of subjects deviate from game-theoretic play: 65.1 percent trust at node 1 of the games, and 58.9 of those who were trusted reciprocate at node 2. On the other hand, subjects do not uniformly choose the evolutionary psychology outcome, though the majority of them do. Clearly the overall results indicate that subjects sometimes do, and sometimes don't, trust and reciprocate. Further analysis gives some insight into when greater trust and reciprocity might be expected. The first Table shows that, for most treatment cells, when icons with facial characteristics are present, subjects are more likely to take an initial trusting move than in the sessions with no icons or non-facial icons, consistent with Prediction 3B above. This is tested more explicitly below.

Game 1 is worth some special attention, since this is the very first game that subjects play, and so less affected by previous history. (Recall that all subjects play the first six games in the same order, after which subjects play as many as 30 games, randomly selected from a set that includes the first six.). About half of the observations for Game 1 shown in Table 1 are from the first move of the experiment.<sup>7</sup> Here we see the clearest indication that the presence of a facial icon makes a difference in the subjects' propensity to trust. In addition, it is in this game that we observe the strongest differences among facial types.

Overall trusting behavior is greatest in Game 2, though it is not clear why this occurs. (It is possible that greater cooperation in this game is due in part to the history of play leading up to it.) There is some variation in trusting behavior by facial pairing. These patterns are analyzed in more detail below, where we also incorporate game history into the analysis.

<Table 1 About Here>

Table 2 contains data on decisions for the second node of the game, contingent on a trusting move being taken. For example, subjects who were assigned the “devious” icon in Game 1 reciprocated in 57.1 percent of the decisions; subjects who faced a “devious” counterpart reciprocated in 35.3 percent of the decisions. In general we see that trust is reciprocated by the second mover, though there appears to be less reciprocity in sessions with non-facial icons or no icons. However, note that in games 2 and 3 this second decision is also a trust decision, since the first mover has a tempting choice at the final node that would give Player 2 a low payoff. Once more we see some variation by facial pairing and across games. (These results should be interpreted with caution, as the number of decisions in some cells is very small, particularly in game 3.)

<Table 2 About Here>

Table 3 shows subjects' equal-split moves at the last stage of the game. In game 1, the two choices are identical, so all subjects who reach the third node achieve an equal split. Consequently these data are omitted from the table. For games 2 and 3, we see that more

---

<sup>7</sup> We performed the analysis on the 51 observations from the first period of play only, and found the same qualitative results as those presented here for all Game 1 observations (106). In the sessions with facial icons, subjects trusted in 69.4 percent of the decisions at node 1; in the sessions without facial icons the corresponding percentage is 33.3. ( $\chi^2=5.7$ ,  $p=.017$ ). At the second node, the pattern of play was similar, though not statistically significantly so because of the small number of observations. In 72 percent of the decisions with facial icons, subjects trusted; for the no-face sessions the percentage was 40, though this represented only two decisions. ( $\chi^2=1.92$ ,  $p=.166$ ).

than half of subjects generally choose an equal split. For example, in game 2, subjects with “devious” icons choose the equal split in 73.1 of the decisions; subjects who face “devious” counterparts choose the equal split in 68.8 percent of the decisions. In the sessions without facial icons, subjects are less likely to choose an equal split. Again, results across games or cells should be interpreted with caution due to the small number of observations.

<Table 3 About Here>

From the tables above, it appears that there are large differences in both trusting and reciprocating when subjects view an icon with facial characteristics versus when they see an icon with no facial content (or no icon at all). Table 4a - c pools the data for all facial-icon cells and all non-facial-icon cells for each game to test Prediction 3A more directly. In Table 4a, for example, we see that subjects trusted in 66.2 percent of decisions in sessions with facial icons, and only 34.4 percent of decisions in sessions without facial icons. (Data for no trust and no reciprocity are suppressed in the table.) The  $\chi^2$ - tests are reported at the bottom of each column. For example, the statistic at the bottom of the first column for Table 4a is for the 2x2 test of face/no face by trust/no trust, and is 9.22, and is statistically significant at conventional levels. The second column tests face/no face by reciprocity/no reciprocity. Except for Table 4c, where the number of observations is small, we see clear statistical differences between decisions where facial icons are present and sessions where they are not. These tests provide strong support for Prediction 3A: The presence of a facial icon alone significantly (and substantially) increases both trust and reciprocity.

<Table 4 about here>

### *Multivariate Models.*

In this section we present Probit regression analysis of the probability of making a trusting or reciprocating decision using data from all three games above. These regressions allow us to test Predictions 3A-C while taking the history of play into account.. Table 5 contains estimates for the probability of trusting or reciprocating as a function of the characteristics of the experimental sessions. The first equation for each model predicts the "first mover" choosing to "trust"; the dependent variable is 1 if the subject trusts by passing at node 1 of the games above. The second equation predicts the “second mover” choosing to “reciprocate”; the dependent variable is 1 if the subject reciprocates by passing at node 2 of the game. For Model 1, independent variables include: Face, a dummy variable equal to 1 if a facial icon is present; and Memory, a variable that measures the a subject’s history.

The variable consists of the number of periods since the subject experienced a low-payoff outcome that was not his choice. In most cases this meant the subject was treated as a “sucker”—that is the subject trusted, but his counterpart did not reciprocate. For Model 2, we add the variable *Icon*, a dummy variable equal to 1 if a nonfacial icon is present. The coefficients for Model 1 show strong evidence of an increase in trusting and reciprocating behavior when there is a facial icon present. The coefficient on *Memory* is positive, but not significant at traditional levels. These results provide additional support for Prediction 3A. In the equations for Model 2, we see that adding the variable *Icon* has little effect on the other coefficients, and that its coefficient is positive but insignificant. This result provides some support for 3B.

<Table 5 about here>

Table 6 presents similar estimates for Prediction 3C. Here the dependent variables are the same as the previous table. The sample used for these estimates includes only those sessions where facial icons were present. Independent variables include *Memory*; a dummy variable equal to 1 if the counterpart of the decision-maker was assigned a Happy or Sad icon; and a dummy variable equal to 1 if the counterpart of the decision-maker was assigned a Devious or Angry icon. (No intercept is included since it would be collinear with Happy/Sad and Devious/Angry.) To test Prediction 3C we need to know if the coefficients on Happy/Sad and Devious/Angry are significantly different. A look at the estimates for the Trust equation indicates no difference: counterparts with a Happy or Sad icon are not trusted more than those with a Devious/Angry icon. However, in the Reciprocity equation, we see that the trusting move of someone with a Happy/Sad icon is more likely to be reciprocated at the second node of the game. We again see some indication that facial characteristics matter. Prediction 3C receives some support, but the results are mixed.

<Table 6 about here>

To further disentangle the effects of facial characteristics on the behavior of experimental subjects, in Table 7 we present additional Trust and Reciprocity regressions which include two variables designed to capture the effects of facial characteristics of both the decision-makers and their counterparts – both equal to 1 for the friendlier version of the characteristic. Here again we restrict the sample to include only observations from sessions with facial icons. The intercept variable is excluded.

Theory gives us no clear guidance in developing hypotheses about the effect on a decision-maker of his own icon. Game theory tells us that subjects base their strategy

choice on the expectations of others' strategy choice. But carry that reasoning one step further: the strategy choice should also depend on what the decision maker thinks his counterpart will infer from his icon signal. Thus decision makers with friendly or nonthreatening icons might be more likely to trust, expecting their icon-type to elicit reciprocity. The results show positive, significant effects for the decision-maker's own smile in the Trust equation, and for the counterpart's eyebrows in the Reciprocity equation. Thus "smilers" are more likely to trust, and expressions with nonthreatening eyebrows are more likely to elicit reciprocity.<sup>8</sup>

<Table 7 about here>

*Discussion.* We have presented analysis of experimental data on the effect of facial icons on trust and reciprocity. Our primary finding is that in the presence of facial icons there is both more trust and more reciprocity than when subjects are presented with nonfacial icons, or when no icons are present. We observe behavior that is more consistent with the predictions of game theory when there are no icons, or when icons do not represent facial expressions. Icons without facial content have no effect on trust or reciprocity. The introduction of the icons increases trust and reciprocity. Subjects who face Happy or Sad icons are more likely to reciprocate a trusting move. In addition we observe that a smiling face is more likely to trust. Our interpretation of this result is that a subject who knows he is signaling a smile anticipates that his counterpart will be more likely to reciprocate a trust move by a smiling face. The subject's own icon is a relatively prominent component of the experiment, and subjects perhaps consider that there is benefit to reinforcing a reputation for play that is consistent with the icon. On the other hand, the counterpart's icon appears to have a weaker effect on the subject's decision whether to trust. This is not surprising, as the counterpart's icon is a weaker signal (or has less information value). The subject faces as many as 7 different counterparts in the course of any given session, with two different types. An icon is a signal of a type – but there are always several players with that type. Thus the decision maker is uncertain about the identity of his counterpart.

## Conclusion

These experiments have not established the foundation and source of reputation. However, what is clear from these data is that subjects do use reputational cues. Those

---

<sup>8</sup> Ideally we would like also to test for interaction effects among the facial-characteristics variables. Unfortunately this awaits the collection of additional data.

cues come directly from simple elements of facial characteristics. When facial characteristics are absent, subjects exhibit little willingness to “trust” their partner. When those characteristics are present, then trusting behavior is plentiful.

The fact that there is such a dichotomy between the presence and absence of simple facial cues poses a serious challenge for game theory. On the one hand, game theoretic predictions do quite well in the absence of reputational priors. This is encouraging because it tells us that game theoretic models appear to be on the right track. On the other hand, those same predictions do miserably when reputational priors are randomly generated. What is especially worthy of note is that only icons with facial characteristics generate reputations. This implies that game theory, which nominally is about the strategic interaction of human beings, needs to pay attention to human beings. As it now stands, for some forms of two-person interaction, standard game theoretic models appear to be a model of autism -- strategic interaction that only looks inward and fails to acknowledge the presence and active participation of others.

These experiments buttress the findings of many others. Simple facial schematics do embody affective content. Our results also point out that these schematics embody social content as well. They are used, and used effectively, to generate trust and reciprocity. This finding moves beyond a simple accounting of emotional content to point out that the inferences drawn by subjects have behavioral characteristics.

Generally these results show that political advisors understand folk psychology. Simple cues can trigger reputational priors and those priors have real content. They result in more or less cooperation in settings in which cooperation is costly. Whether the “trigger” is due to standard concepts of stereotyping, attribution or evolutionary psychology remains an open question. Nonetheless these findings are intriguing and call for additional work.

## Bibliography

- Aronoff, Joel, Andrew M. Barclay and Linda A. Stevenson. 1988. "The Recognition of Threatening Facial Stimuli" *Journal of Personality and Social Psychology*. 54: 647-655.
- Aronoff, Joel, Barbara A. Woike and Lester M. Hyman. 1992. "Which Are the Stimuli in Facial Displays of Anger and Happiness? Configurational Bases of Emotion Recognition." *Journal of Personality and Social Psychology*. 62: 1050-1066.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Baron-Cohen, Simon. 1995. *Mindblindness: An Essay on Autism and Theory of Mind*. Boston: MIT Press.
- Berg, Joyce; Dickhaut, John W.; McCabe, Kevin A. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*; v10 n1, July, pp. 122-42.
- Budesheim, T. L. and S. J. DePaola. 1994. "Beauty or the beast? The effects of appearance, personality and issue information on evaluations of political candidates." *Personality and Social Psychology Bulletin*. 20: 339-348.
- Camerer, Colin, 1998. "Simple Bargaining and Social Utility: Dictator, Ultimatum and Trust Games." Chapter 3 in *Experiments in Strategic Interaction*, forthcoming.
- Cherulnik, P. D. , L. C. Turns and S. K. Wilderman. 1990. "Physical appearance and leadership: Exploring the role of appearance-based attribution in leadership emergence." *Journal of Applied Social Psychology*. 20: 1530-1539.
- Cosmides, Leda and John Tooby. 1992. "Cognitive Adaptations for Social Exchange." In Barkow, Jerome H., Cosmides and Tooby, eds., *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. 1992. New York: Oxford University Press.
- Eckel, Catherine C. and Philip Grossman, "Altruism in Anonymous Dictator Games." *Games and Economic Behavior* 16:181-191, 1996.
- Eckel, Catherine C. and Philip Grossman, "Are Women Less Selfish Than Men?: Evidence from Dictator Games." Forthcoming, *The Economic Journal*, May, 1998.
- Darwin (1998) -- *The Expression of the Emotions in Man and Animals* (Series in Affective Science) by Charles Darwin. Paul Ekman (Editor) Oxford University Press.
- Ekman, Paul (1972) -- *Emotion in the Human Face : Guide-Lines for Research and an Integration of Findings*. (Pergamon general psychology series ; PGPS-11) New York: Pergamon Press.
- Ekman, Paul. 1982. *Emotion in the human face*, 2nd ed. (Studies in emotion and social interaction) New York: Cambridge University Press.
- Ekman, Paul. 1997. *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (Series in affective Science) New York : Oxford University Press.
- Ekman, P. and W. V. Friesen. 1978. *The Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., W. V. Friesen and M. O'Sullivan. 1998. "Smiles when lying." *Journal of Personality and Social Psychology*. 54: 414-420.



- Fernandezdols, J. M. and M. A. Ruizbelda. 1995. "Are smiles a sign of happiness -- Gold Medal winners at the Olympic games." *Journal of Personality and Social Psychology*. 69: 1113-1119.
- Fehr, Ernst, Georg Kircksteiger, and Arno Reidl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108(2): 437-59.
- Forsythe, R., Horowitz, J. L., Savin, N. E. and Sefton M. (1994). 'Fairness in simple bargaining experiments.' *Games and Economic Behavior*, vol. 6, pp. 347-369.
- Fridlund, Alan J. 1994. *Human Facial Expression: An Evolutionary View*. San Diego: Academic Press.
- Fundenberg, D. and E. Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica*. 54: 533-544.
- Gopnik, A. 1993. Mindblindness. Unpublished essay, University of California, Berkeley. Cited in Baron-Cohen (1995).
- Hamilton, W. D. 1964. "The Genetical Evolution of Social Behaviour (I and II)." *Journal of Theoretical Biology* 7: 1-16, 17-52.
- Hoffman, E., McCabe, K., Shachat, K. and Smith, V. (1994). 'Preference, property rights and anonymity in bargaining games.' *Games and Economic Behavior*, vol. 7, pp. 346-80.
- Hoffman, E., McCabe, K. and Smith, V. (1996). 'Social distance and other-regarding behavior in dictator games.' *American Economic Review*, vol. 86, pp. 653-60.
- Hoffman, E., McCabe, K. and Smith, V. (forthcoming) "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology." *Economic Inquiry*.
- Johnson, Mark H., Suzanne Dziurawiec, Haydn Ellis and John Morton. 1991. "Newborns preferential tracking of face-like stimuli and its subsequent decline." *Cognition*. 40: 1-19.
- Katsikitis, Mary. 1997. "The classification of facial expressions of emotion: a multidimensional scaling approach." *Perception*. 26: 613-626.
- Kreps, D.M., J.D. Roberts, P. Milgrom, and R. Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma." *Journal of Economic Theory*. 27: 245-252.
- Le Doux, Joseph 1996 *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.
- Leonard, C. M., K. K. S. Voeller and J. M. Kuldau. 1991. "When's a smile a smile? Or how to detect a message by digitizing the signal." *Psychological Science*. 2: 166-172.
- MacDonald, P.M, S.W. Kirkpatrick and L. A. Sullivan. 1996. "Schematic Drawings of Facial Expressions for Emotion Recognition and Interpretation by Preschool-Aged Children." *Genetic, Social and General Psychology Monographs*. 122: 375-388.
- Masters, R. D. and D. G. Sullivan. 1989. "Facial Displays and Political Leadership in France." *Behavioral Processes*. 19: 1-30.
- McKelvie, Stuart J. 1973. "The meaningfulness and meaning of schematic faces." *Perception and Psychophysics* 14 (2): 343-348.
- Ostrom, Elinor. 1998. "A Behavioral Approach to the Rational Choice Theory of Collective Action." *American Political Science Review*. 92: 1-22.
- Pinker, Steven. 1994. *The Language Instinct*. New York: Harper-Collins.

- Pinker, Steven. 1997. *How the Mind Works*. New York: W. W. Norton & Co.
- Russell, J. A. 1993. "Forced-Choice Response Format in the Study of Facial Expression." *Motivation and Emotion*. 17: 41-51.
- Sorce, J. F., R. N. Emde, J. J. Campos and M. D. Klennert. 1985. "Maternal emotional signaling: Its effects on the visual cliff behavior of 1-year-olds." *Developmental Psychology*. 21: 195-200.
- Sullivan, Denis G. and Roger D. Masters. 1988. "'Happy Warriors': Leaders' Facial Displays, Viewers' Emotions, and Political Support." *American Journal of Political Science*. 32: 345-368.
- Tajfel, Henri and J. Turner, "An Integrative theory of intergroup conflict," pp. 1-10. In J. Turner, R. Haslam, A. Haslam, and S. Worchel (eds.) *The Social Psychology of Intergroup Relations*. CA: Brooks/Cole (1979).
- Tooby, John and Leda Cosmides. 1992. "The Psychological Foundations of Culture." In Jerome H. Barkow, Leda Cosmides and John Tooby (eds.). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Cambridge: Oxford University Press, pp. 19-136.
- Trivers, R. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46: 35-57.
- Turner, John, "Social Categorization and Social Discrimination in the Minimal Group Paradigm." in Tajfel, Henri. *Differentiation Between Social Groups: Studies in Social Psychology of Intergroup Relations*. London: Academic Press Inc. (1975).
- Wright, Robert. 1994. *The Moral Animal: Why We Are the Way We Are: The new Science of Evolutionary Psychology*. New York: Vintage Books.
- Yamada, Hiroshi. 1993. "Visual Information for Categorizing Facial Expression of Emotion." *Applied Cognitive Psychology* 7: 257-270.
- Zebrowitz, Leslie A. 1997. *Reading Faces: Window to the Soul?* Boulder, CO.: Westview Press.

Figure 1  
Sequential Game One

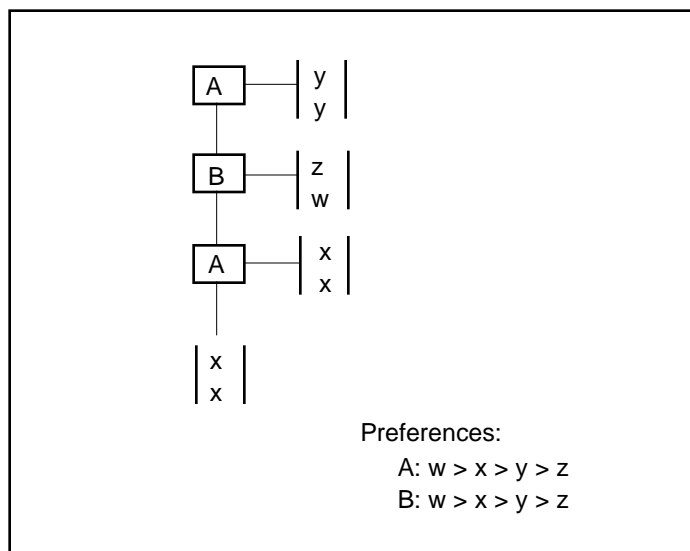


Figure 2  
Sequential Game Two

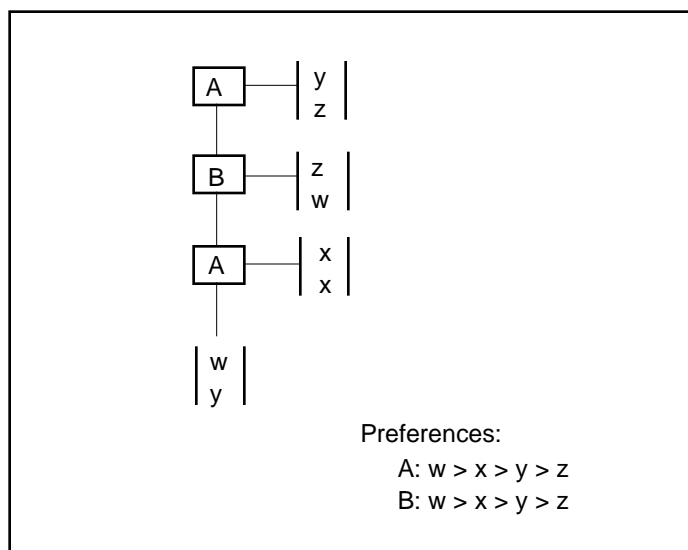


Figure 3  
Icons Used in Experiment 1

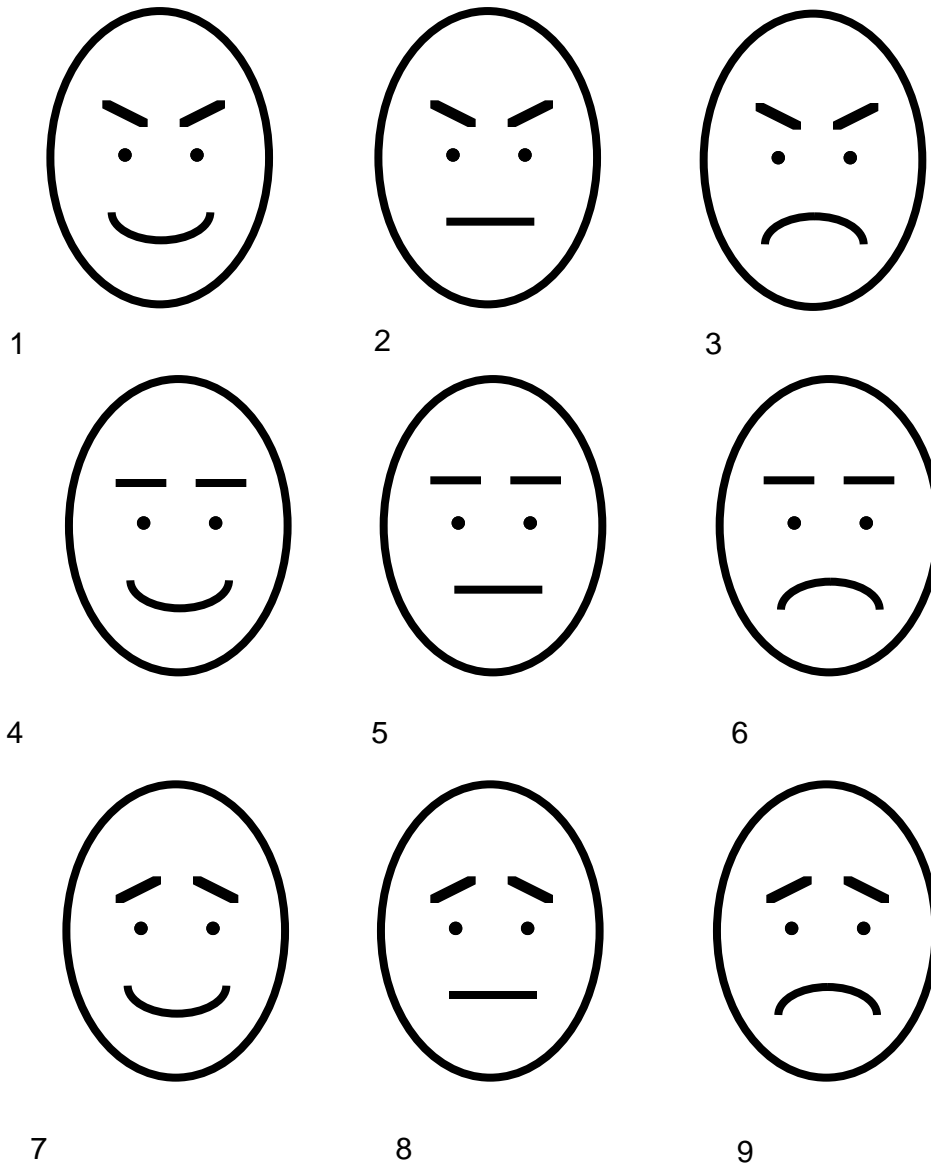


Figure 4  
Mean Response Across 7-point Semantic Differential Scale by Icon

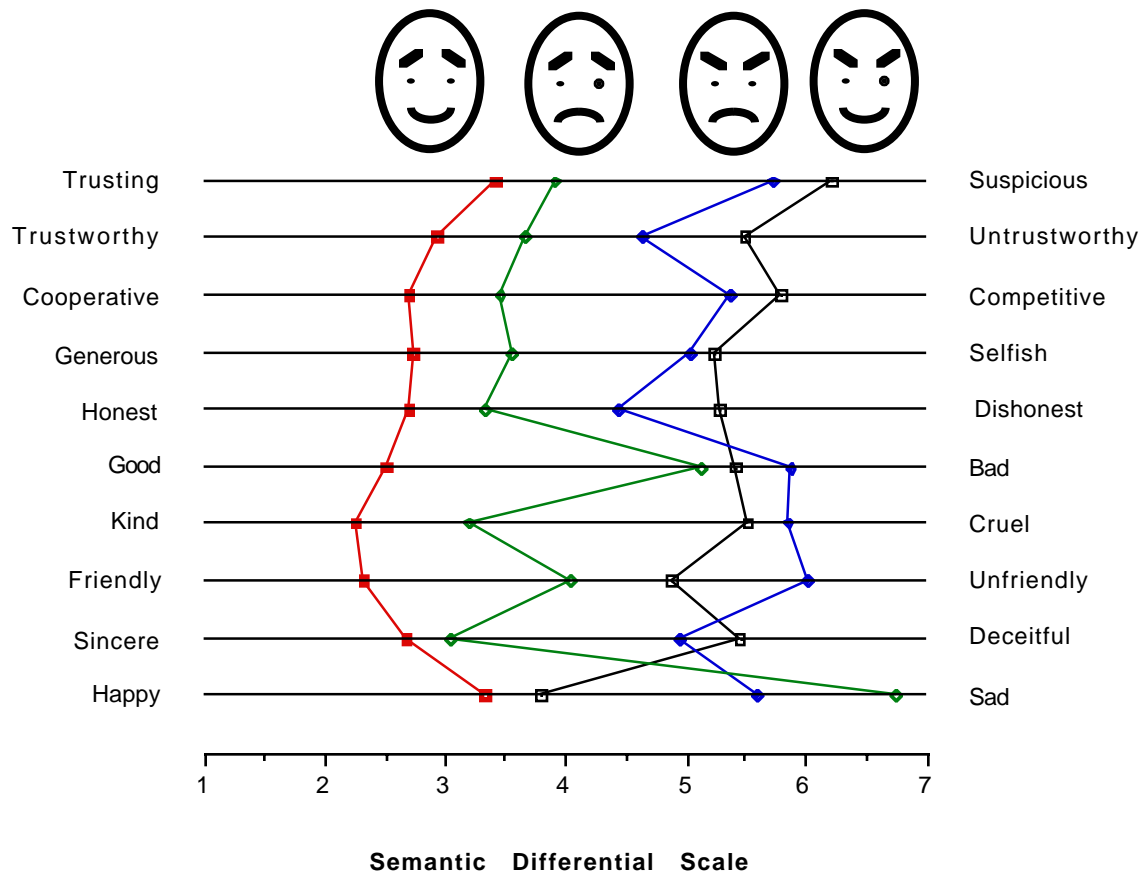


Figure 5  
Two-person Sequential Games and Payoffs Observed by Subjects

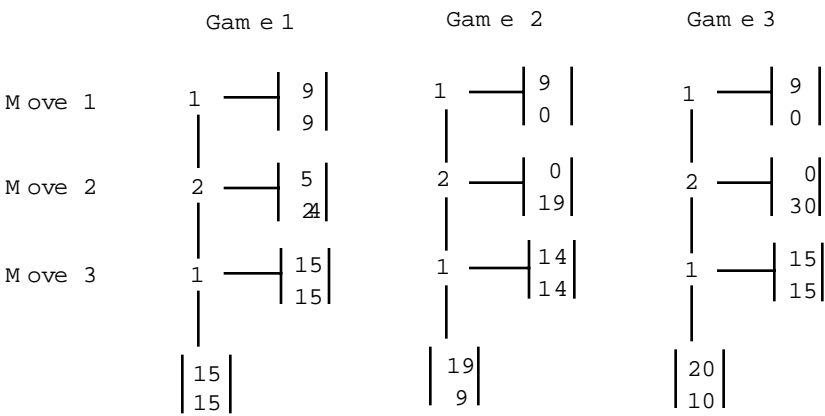


Figure 6  
Icon Types Used in Experiment 2

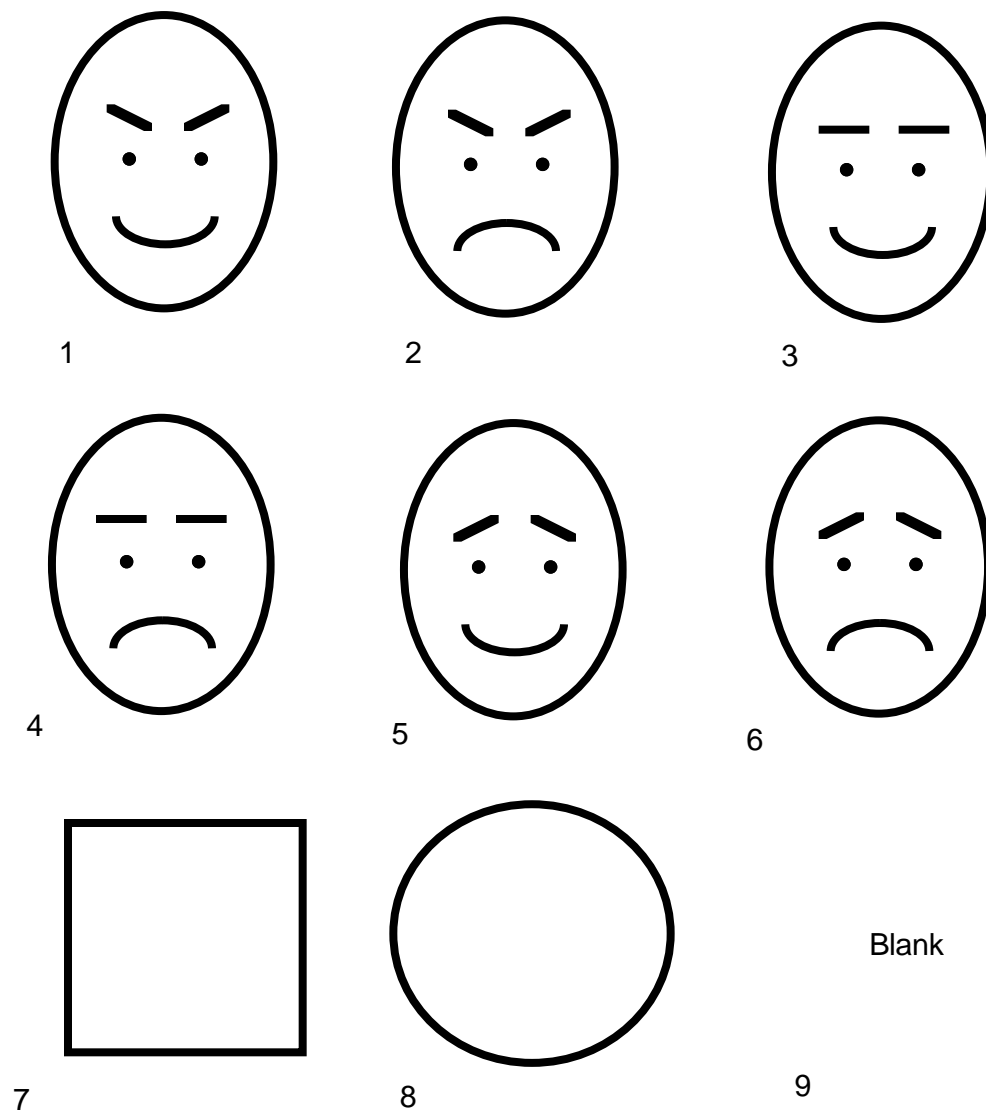




Table 1: Percentage of First Movers Selecting a “Trusting” Move  
by Own and Other's Icon, Games 1-3  
(Frequencies in Parentheses)




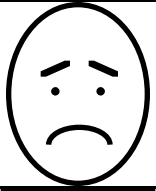
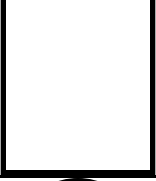
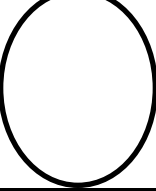
Icon	Game 1		Game 2		Game 3	
	Own Face	Other Face	Own Face	Other Face	Own Face	Other Face
	77.3 (17)	72.4 (21)	72.7 (16)	74.3 (26)	61.5 (8)	68.4 (13)
	44.4 (4)	46.2 (6)	85.7 (12)	89.5 (17)	100.0 (2)	100.0 (5)
	77.8 (21)	69.6 (16)	94.6 (35)	83.3 (25)	83.3 (10)	66.7 (6)
	43.7 (7)	66.7 (6)	62.5 (10)	100.0 (5)	50.0 (4)	-- (0)
	28.6 (2)	37.5 (3)	75.0 (6)	55.6 (5)	66.7 (2)	33.3 (1)
	45.5 (5)	40.0 (8)	66.7 (6)	87.5 (7)	25.0 (1)	50.0 (2)
(Blank)	28.6 (4)	28.6 (4)	53.3 (8)	53.3 (8)	40.0 (2)	40.0 (2)

Table 2: Percentage of Second Movers Selecting a “Reciprocity” Move  
by Own and Other Icon, Games 1-3  
(Frequencies in Parentheses)





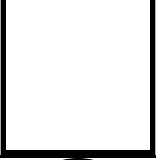
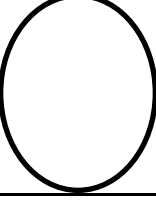
Icon	Game 1		Game 2		Game 3	
	Own Face	Other Face	Own Face	Other Face	Own Face	Other Face
	57.1 (12)	35.3 (6)	73.1 (19)	68.8 (11)	53.9 (7)	50.0 (4)
	83.3 (5)	75.0 (3)	76.5 (13)	75.0 (9)	80.0 (4)	50.0 (1)
	68.8 (11)	81.0 (17)	80.0 (20)	80.0 (28)	50.0 (3)	70.0 (7)
	50.0 (3)	71.4 (5)	20.0 (1)	50.0 (5)	-- --	50.0 (2)
	33.3 (1)	100.0 (2)	40.0 (2)	50.0 (3)	0.0 (0)	100.0 (2)
	25.0 (1)	0.0 (0)	71.4 (5)	66.7 (4)	100.0 (2)	0.0 (0)
(Blank)	25.0 (1)	25.0 (1)	25.0 (2)	25.0 (2)	50.0 (1)	50.0 (1)

Table 3: Percentage of First Movers Selecting an Equal Split  
at the Last Move by Own and Other Icon, Games 2-3  
(Frequencies in Parentheses)

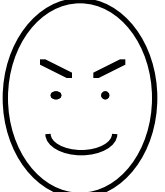
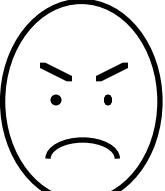
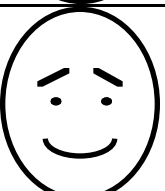
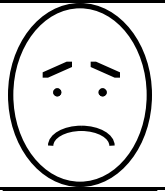
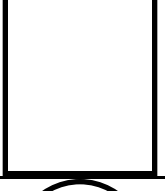
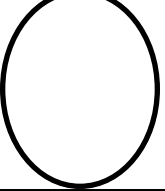
Icon	Game 2 Own Face	Game 2 Other Face	Game 3 Own Face	Game 3 Other Face
	73.1 (19)	68.8 (11)	53.9 (7)	50.0 (4)
	76.5 (13)	75.0 (9)	80.0 (4)	50.0 (1)
	80.0 (20)	80.0 (28)	50.0 (3)	70.0 (7)
	20.0 (1)	50.0 (5)	--	50.0 (2)
	40.0 (2)	50.0 (3)	0.0 (0)	100.0 (2)
	71.4 (5)	66.7 (4)	100 (2)	0 (0)
(Blank)	25.0 (2)	25.0 (2)	50.0 (1)	50.0 (1)

Table 4  
Trusting and Reciprocal Behavior Using Pooled Data  
a. Game 1

	Trust	Reciprocity
Face	66.2 (49)	63.3 (31)
No Face	34.4 (1)	27.3 (3)
	$\chi^2=9.22, p=.002$	$\chi^2=4.74, p=.03$

b. Game 2

	Trust	Reciprocity
Face	82.0 (73)	72.6 (53)
No Face	62.5 (20)	45.0 (9)
	$\chi^2=5.04, p=.025$	$\chi^2=5.38, p=.020$

c. Game 3

	Trust	Reciprocity
Face	68.6 (24)	58.3 (14)
No Face	41.7 (5)	60.0 (3)
	$\chi^2=2.74, p=.098$	$\chi^2=.005, p=.95$

Table 5  
Probit Estimates for Predictions A and B  
(Standard Errors in Parentheses)

	<b>Model 1</b>		<b>Model 2</b>	
	Trust	Reciprocity	Trust	Reciprocity
<b>Intercept</b>	.163 (.160)	-.277 (.229)	-.341 (.232)	-.643 (.367)
<b>Face (1=yes)</b>	.668** (.175)	.640** (.237)	.840** (.238)	.999** (.369)
<b>Icon/No Icon (1=yes)</b>	--	--	.310 (.292)	.577 (.444)
<b>Memory</b>	.018 (.013)	.011 (.015)	.019 (.013)	.012 (.015)
<b>Log Likelihood</b>	-165.58	-116.64	165.01	115.78

\*\*p < .01, \*p<.05

Table 6  
Probit Estimates for Prediction C  
(Standard Errors in Parentheses)

	<b>Trust</b>	<b>Reciprocity</b>
<b>Happy/Sad</b> <b>(1=yes)</b>	.558** (.190)	.517** (.189)
<b>Devious/Angry</b> <b>(1=yes)</b>	.537** (.157)	.087 (.201)
<b>Memory</b>	.013 (.014)	.015 (.016)
<b>Log Likelihood</b>	-113.610	-90.036

\*\*p < .01, \*p<.05

Table 7  
Probit Estimates for Mouth and Eyebrow Shapes  
(Standard Errors in Parentheses)

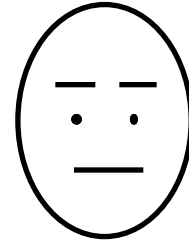
	Trust	Reciprocity
Own Eyebrow (1= / \)	.230 (.182)	-.070 (.218)
Other Eyebrow (1= / \)	-.087 (.204)	.447* (.204)
Own Mouth (1=smile)	.753** (.208)	-.006 (.243)
Other Mouth (1=smile)	-.112 (.207)	.126 (.244)
Memory	.013 (.014)	.016 (.015)
Log likelihood	-108.66	-89.93

\*\*p < .01, \*p<.05

## Appendix 1

First page of the survey instrument used in experiment 1. The sample icon is of a neutral image. Eight other icons were also used.

To the right you are presented with an icon. This icon is supposed to represent a type of person. I am very interested in what you think the icon means. Below I provide pairs of words that have opposite meanings. On each line please fill in the circle with whichever word you think fits the icon.



How Close is the Word to the Icon?

	Very	Somewhat	Slightly	Neither	Slightly	Somewhat	Very	
good	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	bad
weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	strong
excitable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	calm
cruel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	kind
attractive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unattractive
suspicious	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	trusting
pleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unpleasant
fragile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	tough
active	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	passive
unfriendly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	friendly
competitive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	cooperative
vengeful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	forgiving
honest	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	dishonest
selfish	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	generous
trustworthy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	untrustworthy
inconsiderate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	considerate
sincere	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	deceitful
malevolent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	benevolent
submissive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	domineering
sad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	happy
male	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	female
scheming	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	forthright
content	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	frustrated
insecure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	secure
amiable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	hostile



## Appendix 2

### Instruction Set

Screen 1

**In this experiment you will participate in several two person decision problems. At each decision you will be randomly paired with another individual in this room: your counterpart.**

**The joint decisions made by you and your counterpart will determine how much money you will earn for this decision problem.**

**Your earnings for this decision will be paid to you in cash at the end of this experiment. I will not tell anyone else your earnings and I ask you not to discuss your earnings.**

**Click OK when you are ready to continue.**

OK

---

Screen 2

**You will not be paid for every decision in the experiment. You will make many decisions with the other participants in this experiment.**

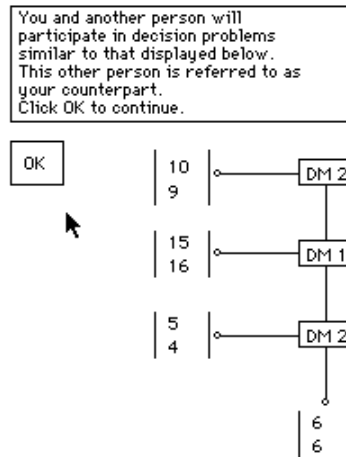
**At the conclusion of the experiment, ONE of the decisions will be randomly selected. You will be paid for that decision.**

**On the sheet of paper I have provided, please record your potential earnings for each decision. This will help you keep track of what you earn at the end.**

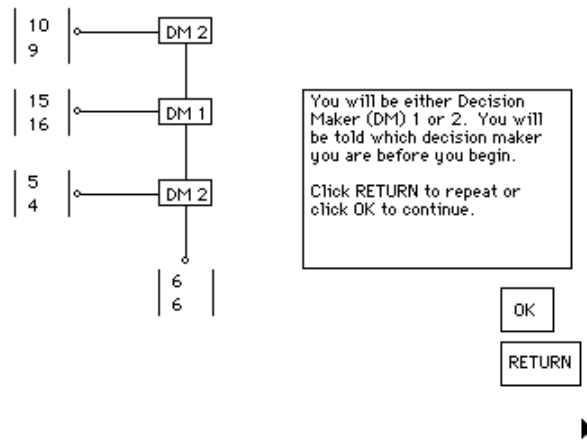
**Click OK when you are ready to continue.**

OK

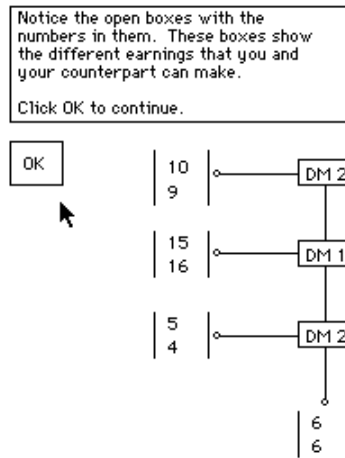
Screen 3



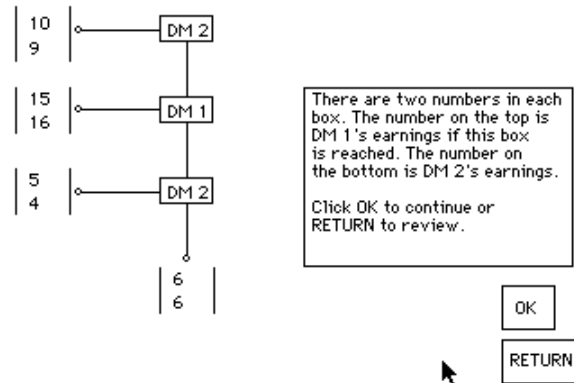
Screen 4



Screen 5

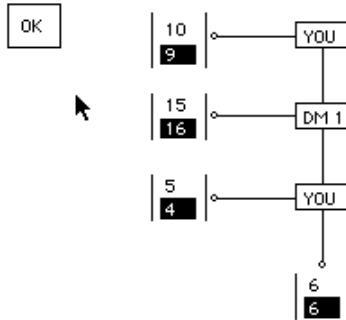


Screen 6

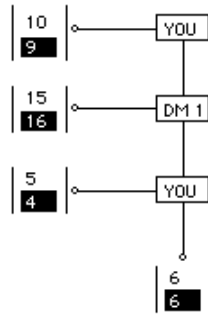


Screen 7

In this example suppose you are DM 2.  
In each box your possible earnings  
are highlighted.  
Notice how the earnings differ in  
each box.  
Click OK to continue.



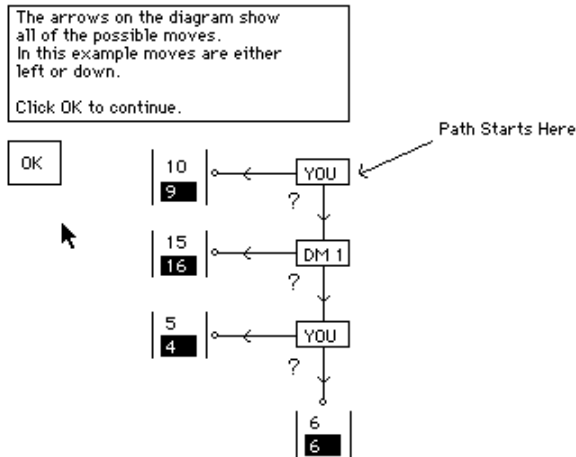
Screen 8



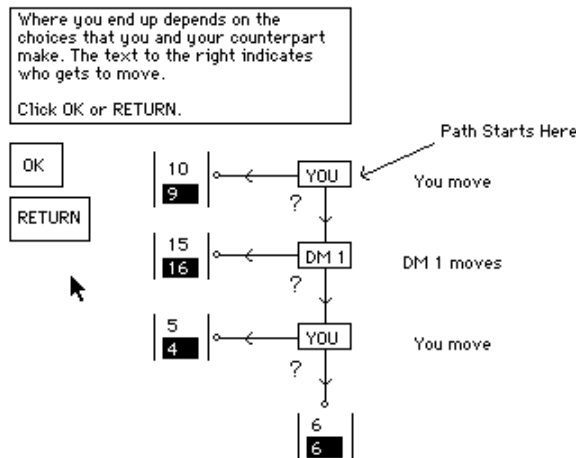
You and your counterpart will  
jointly determine a path  
through the diagram to an  
earnings box. A path starts at  
the top of the diagram. A move  
is a choice of direction on  
the diagram.  
Click OK to continue  
or RETURN to review.

OK  
RETURN

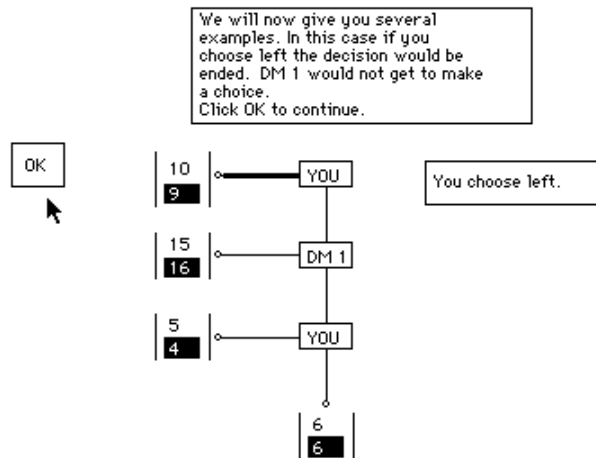
Screen 9



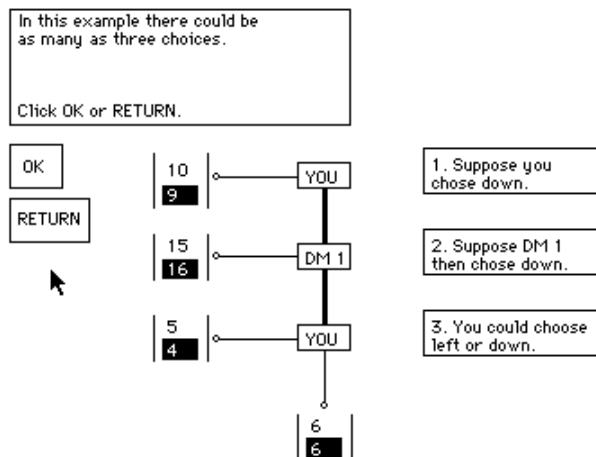
Screen 10



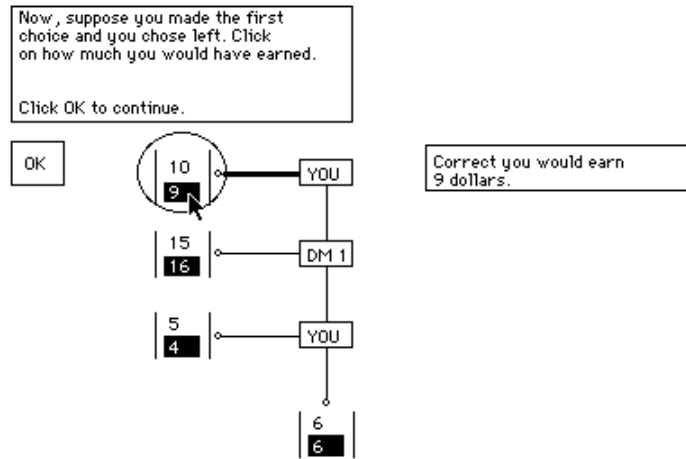
Screen 11



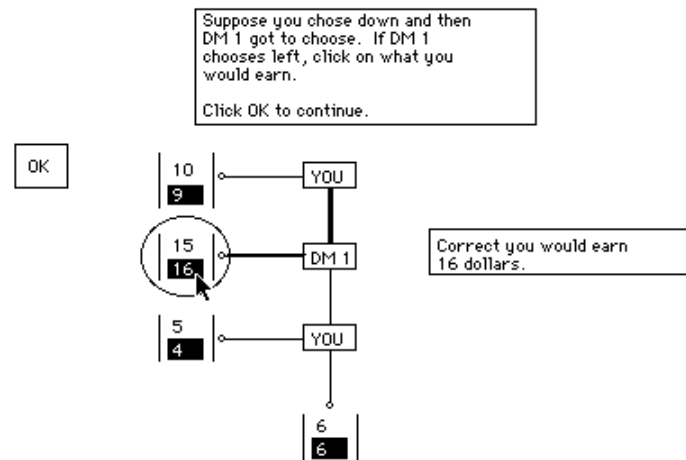
Screen 12



Screen 13

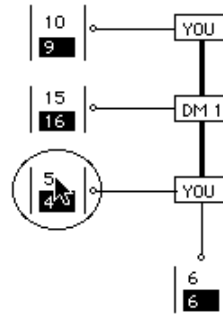


Screen 14



Screen 15

Now, suppose that you chose down,  
then DM 1 chose down. Click on  
what you would earn if you  
chose left.

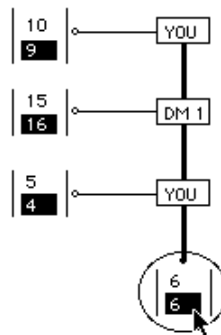


Incorrect, please try again.  
Your earnings are highlighted.

Screen 16

Finally, suppose that you chose down,  
then DM 1 chose down. Click on  
what you would earn if you  
chose down.  
Click OK or RETURN.

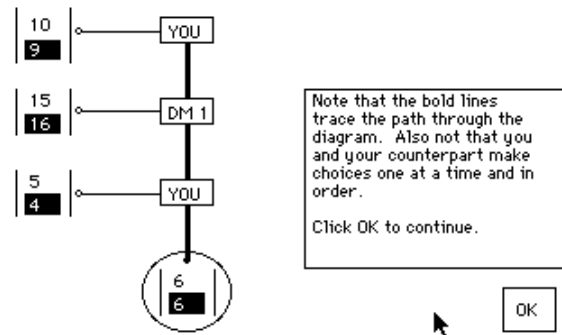
OK  
RETURN



Correct you would earn  
6 dollars.

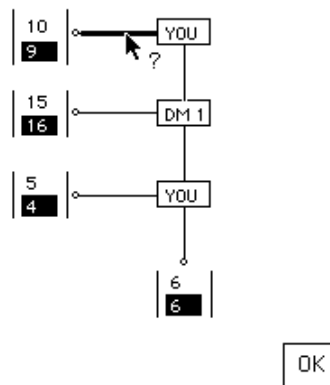


Screen 17

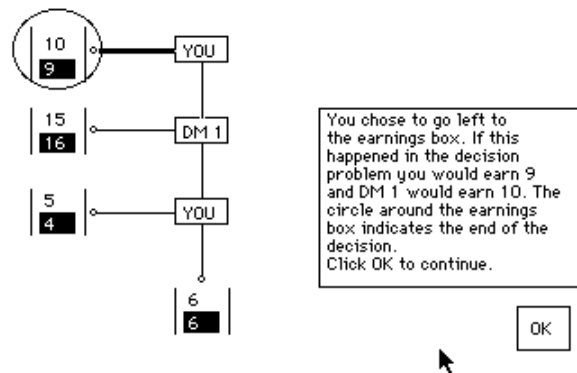


Screen 18

Here is an example of moving through the diagram. The blinking lines and the question mark indicate it is your turn to make a choice. Click on a blinking line to make your choice and then click OK to continue.

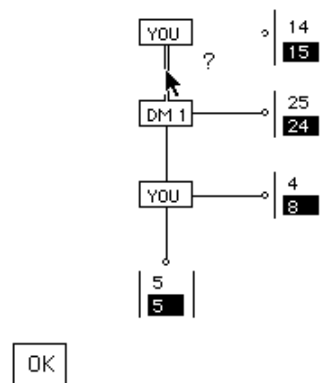


Screen 19

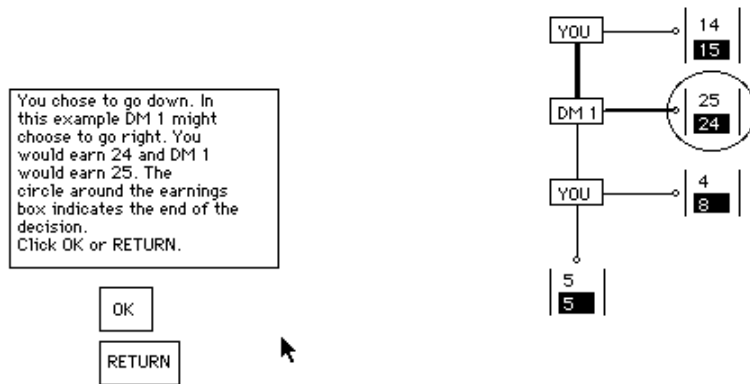


Screen 20

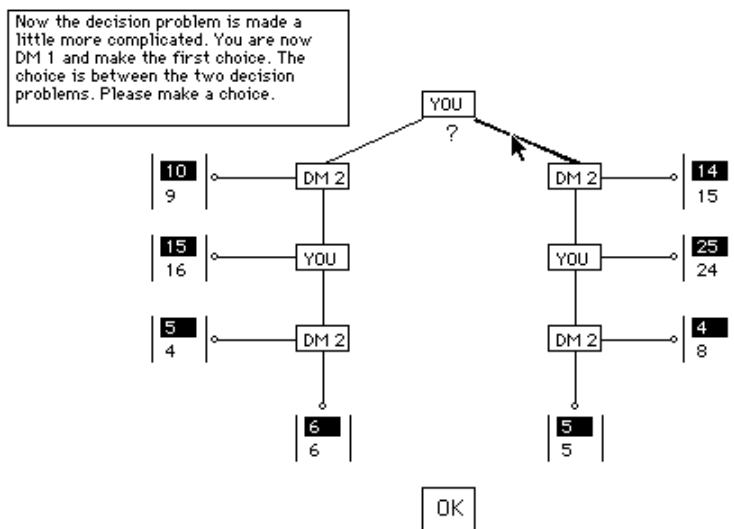
The same principle holds if the decision problem looks like the following. Again assume you are DM 2. Please make a choice by clicking on a blinking line.



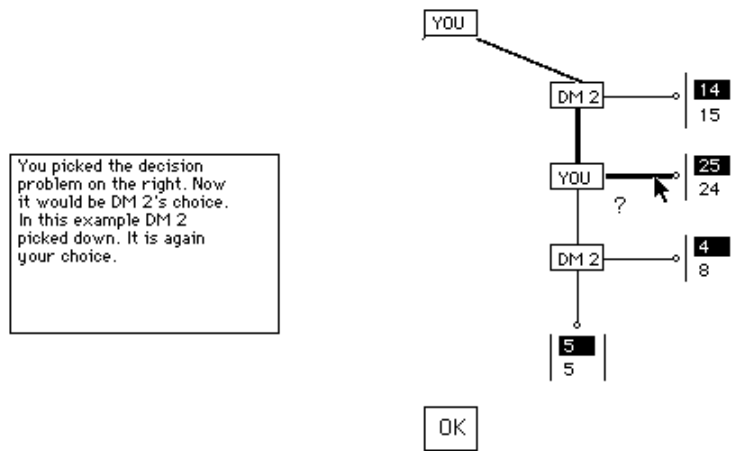
Screen 21



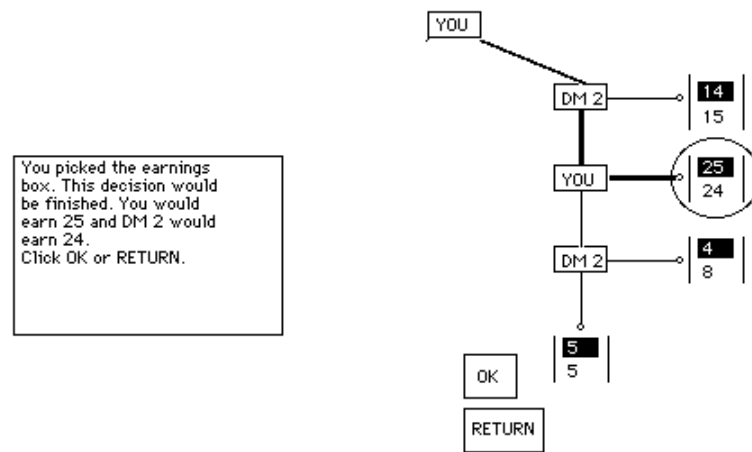
Screen 22



Screen 23



Screen 24



Screen 25

You are about ready to begin. You will make 30 different decisions. However, you will only be paid for one of the decisions that you and your counterpart make. At the end of all of your decisions, you will get to randomly pick one of decisions for which you will be paid. You have a sheet of paper and a pencil to mark your earnings from each decision. Please keep track of how much you could make following each decision. If you have any questions, please ask them now. Otherwise click OK to continue.

OK

---

Screen 26

Finally, during this experiment you will be represented by the icon illustrated to the right. This is what your counterpart will see before beginning a decision problem. Likewise you will see the icon for your counterpart.



Click OK to continue or RETURN to review.

OK

RETURN