2/8/99 10:16 AM

The Human Face of Game Theory: Trust and Reciprocity in Sequential Games.

Catherine C. Eckel
Department of Economics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
eckelc@vt.edu

and

Rick K. Wilson
Department of Political Science
Rice University
Houston, TX 77251-1892
rkw@rice.edu

This paper was prepared for Trust Working Group Meeting and sponsored by the Russell Sage Foundation, February 19-20, 1999, New York, NY. Valuable assistance for the classroom experiments conducted at VPI was provided by Fatma Aksal, Robert Bowles, Tony Dziepak, Rob Gilles, Cathleen Johnson, Mark McLeod and Randall Verbrugge. Jane Sell and Carl Rhodes provided critical assistance in conducting the computer controlled experiments at Texas A&M. Sean Bolks and Laura Delgado helped conduct experiments at Rice University. Finally, we spent some time bending the ears of Betsy Hoffman, Kevin McCabe and Vernon Smith

Abstract

The Nash equilibrium predictions of standard game theory often fail in experimental tests. While innumerable refinements to Nash equilibrium have been suggested, they too fail in empirical tests. What seems apparent is that humans rely on a rich mixture of contextual cues and signals to determine how to strategize about their own choices and those of their counterparts.

Our research concentrates on a simple social signal -- the human face. In particular we examine the emotional cues registered by the face and investigate the meaning of those cues mean for signaling an actor’s intentions. We explore the impact of facial cues in two sets of experiments. The first concerns human facial expressions. In the second set of experiments we simplify and control the facial cues by substituting schematic line drawings that resemble the ubiquitous "happy face". We focus on the two-branch, sequential bargaining game developed by Hoffman, McCabe and Smith. Our findings suggest that intentions to trust and reciprocate can be signaled by facial expressions.

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

Introduction

In our daily interactions with others we place remarkable trust in complete strangers. When crossing the street we expect that a person behind the steering wheel will yield to pedestrians (although we quickly scan the driver's eyes to make certain they see us); when buying a bottle of wine, we assume that it has been properly stored (although we often return to the same wine shop once we have found an owner who is trustworthy); and we give the keys to the house to a weekly cleaning service (and are shocked and dismayed if our trust is betrayed by theft.)

This idea of trusting complete strangers is antithetical to much of Western social thought. Reaching back to Machiavelli or Hobbes, it is argued that individuals will exploit others who are trusting. Social contractarians like Locke soften this characterization only by noting the importance of a well-defined context for social interaction that is capable of curbing the exploitative nature of humans. Madison argues that the "various and unequal distribution of property" makes mankind "much more likely to vex and oppress each other than to cooperate for their common good." (*Federalist Papers, # 10*). Even Adam Smith, who points to the benefits derived from the "invisible hand" of faceless economic exchange, relies heavily on a context provided by well defined property rights guaranteed by a political entity.[1]

In the second half of the 20th century, there has been a systematic attempt to mathematize the complex relationships between individuals. Game theorists have carried out much of this work, the bulk of which has focused on "non-cooperative" game theory. Such an approach has its roots in Hobbesian models of human nature. The fundamental implication of this work is that in any one-shot game, trusting another player is a risky strategy. A rational actor always chooses a strategy that anticipates the (rational) strategic play of the counterpart. Often the collective outcome of individual rational play is inferior for all actors, leaving unreachable outcomes that are Pareto superior. The problem, however, is that no actor has an incentive to unilaterally change her strategy,

---

[1] Smith's earlier work on "moral sentiments" is quite different in its approach, and explores the basis of trust and reciprocity in human interconnectedness. See Smith, 1998.

and there is no mechanism to enforce cooperative action if all actors simultaneously are given the chance to change their strategies. Games like “Chicken,” “Battle of the Sexes,” and the “Prisoner’s Dilemma” have most clearly presented this problem and been the principle focus for theorists trying to understand cooperative behavior in settings that are modeled using the tools of non-cooperative game theory.

What has been bothersome to theorists is the widespread observation that people cooperate in environments that similar to many non-cooperative games with Pareto-inferior Nash equilibria. Axelrod (1986) tells of tacit truces halting fighting along the front during World War I and Ostrom (1990) writes about Turkish fishers inventing rotation schemes to prevent over-fishing and to reduce conflict over choice fishing spots. Solutions to these problems may involve very complicated social structures. On the other hand social scientists from the disciplines of psychology, economics, sociology, and political science have found many similar patterns of behavior in settings with minimal institutional structure and with a high degree of experimental control. For example, in prisoner's dilemma game experiments, study after study reports a substantial amount of spontaneous cooperation. (See Rapaport and Chammah, 1965, and the survey by Rapoport, 1988. Meta-surveys of an abundant literature on public goods games make a very similar point. See Dawes, 1980; Ledyard, 1995; Sally, 1995).

Much of the more recent theoretical work has turned to repeated interactions among individuals as a way of establishing cooperation. With repeated interaction among a set of agents, the Folk Theorem concludes that any of a large set of outcomes can be supported as equilibria. Dynamic strategies such as tit-for-tat lead to outcomes that appear fully cooperative; in repeated interactions, agents' reputations embody their strategies, and cooperative outcomes are equilibria of the supergame. This analysis implicitly assumes that players have an initial prior about their counterparts, and acquire a great deal of information about their counterparts through observing strategic play.

While this recent theoretical work has produced important insights about the emergence of cooperation, two things are missing. First there is little in these models to address the fact that individuals employ cooperative strategies *with strangers* with whom they expect no future interaction. Second, little attention is paid to the sources of the initial priors that individuals use to assess the reputations of their counterparts.

The high degree of initial cooperation among strangers is a fascinating empirical regularity. This is not to say that all individuals always begin by playing a cooperative strategy, nor is it the case that the *same* individual always begins by playing a cooperative strategy. Individuals form initial expectations about their counterparts that are crucial for triggering the propensity to cooperate. Our view is that humans share a capacity to read one another's intentions. The ability to read intentions has developed along with the mechanism for displaying intentions.

The question then arises of how intentions are signaled and read. Most species of mammals use facial expressions indicate their intentions through the display of emotions such as anger (threat) or fear; humans are little different in this respect. However, in humans, evolution has shaped a species that is even more acutely attuned to social display. At the same time the capacity for language has aided in developing cultural mechanisms that lead to socially understood meanings with respect to social display. (See Boyd and Richerson, 1985; Deacon, 1997; Dunbar, 1997). Both biologically derived pre-attention to social signals and socially-derived meanings for social signals create a foundation for establishing initial reputations, and initial reputation is a foundation for trusting behavior with strangers.

There are many possible sources for social signals. People use all the information available to them in forming expectations about others' play. This includes information about "type" that might be correlated with strategy choice -- such as sex or race -- but also many conscious and unconscious communications through language, facial expression, and body language. In this paper we focus our investigation on one of the more obvious sources of social signaling -- the human face.

Motivation

Game theory has provided a very successful model for the social sciences. In many settings it fares very well as a predictive model -- particularly where there are many actors, their interactions are anonymous, and the underlying institutional mechanisms are incentive compatible (see Ostrom, 1998). However, when engaged in face-to-face bargaining where the actors clearly are not anonymous, and under relatively weak institutional mechanisms, standard game theory is much less successful in predicting behavior.

A large number of empirical papers show that in bargaining situations laboratory subjects often choose strategies that appear to be cooperative or reciprocal, in contrast to the best-response play predicted by game theory. Subjects sometimes fail even to play dominant strategies, where these are available. Many recent papers are centrally concerned with whether and how actors look beyond themselves and to their partners. This other-regarding behavior occurs in many settings where subject split a fixed pie, including ultimatum and dictator games (Camerer, 1998; Eckel and Grossman, 1996, 1998; Forsythe, et al., 1994; Hoffman, et al., 1994). In situations where there are gains to cooperation, such as public goods games (Ledyard, 1995), investment trust games (Berg et al., 1995) and gift exchange experiments (Fehr, et al., 1993), even more other-regarding behavior is observed. These results are neither random nor haphazard; behavior inconsistent with game theoretic predictions is routine and patterned.

The Problem.

Traditional game theory describes rational actors who play best-response to the strategies of others, which yields a Nash equilibrium. Consider two distant cousins who have a common interest in wine, and a complex inheritance from their great-uncle Julio. Uncle Julio's will names Cathy to inherit her choice of one of two tracts of land. It also specifies that Rick have the responsibility for planting and harvesting the land, and producing the appropriate wine. Cathy and Rick must split the resulting product, in a manner agreeable to them both. The hillside tract is suitable for producing Chardonnay, which Cathy loves, while the mountainside tract is larger, but better suited for Cabernet Sauvignon, which she likes, but not as well as Chardonnay. Rick has the viniculture skills to plant and make the wine. He strongly prefers Cabernet, but could make either varietal grow at either site. While he offers to plant a 60/40 combination of Cabernet/Chardonnay on the mountainside tract and a 40/60 combination of Cabernet/Chardonnay on the hillside tract, Cathy knows little about grapevines and cannot determine if he will carry out his promise. Once the wine is harvested, of course, it is too late. She knows that he is a good enough vintner that he won't plant all Cabernet on the hillside tract, because the Chardonnay will thrive and be wonderful. However, she also knows that the Cabernet from the mountainside is likely to be spectacular and he may plant the entire site in those vines and she will not get any Chardonnay.

Traditional game theory would tell Cathy to look no further than herself. If she had Rick's preferences instead of her own, how would she plant the vineyards? Irrespective of a promise, she would plant the mountainside fully with Cabernet, and probably plant the bulk of the hillside with Cabernet. Reasoning thus, she would choose the hillside tract, enjoy her Chardonnay and drink acceptable Cabernet on occasion. But if she could trust Rick to keep his promise, the mountainside tract would be the better choice.

But what kind of world would this be in which a bargaining situation occurs independently of Cathy and Rick having some understanding of one another? The key question is whether (and under what circumstances) might Cathy be willing to "trust" Rick's offer to plant mixed percentages of each grape? If they had known one another for some time, then past behavior might be sufficient to judge the extent to which Rick keeps his word. Or it may be that Rick has dealt with a number of others whom Cathy knows and she can infer something about his reputation based on their comments. Or, upon first meeting him, she makes a quick assessment of his demeanor, physical stance and face and from that reaches a quick judgment. Any such judgment will be fraught with uncertainty and "trusting" will be risky. However, there are real gains from taking such a trust move and having it reciprocated. Perhaps Cathy can do better by trying to read Rick's "intentions."

In order to provide a little more structure to this story and to set the stage for the experimental design elaborated below, consider the following two games. Both are two-person sequential games with perfect and complete information. The first game we characterize as a simple Nash game while the second game also has a Nash equilibrium, but incorporates gains to cooperation.

First consider Game A given in extensive form by Figure 1. Two actors, A and B, face a series of moves through the game. Payoffs are given as dollars and by convention Player 1's payoffs are in the upper position and 2's are in the lower. Player 1 takes the first move, deciding to go left or right. Player 2 then decides whether to quit the game, taking the payoff, or passing the move back to Player 1. Player 1 then has a choice to quit, taking the payoff, or passing the move back to Player 2. If play reaches the final node, then Player 2 choose between either of two payoff boxes. Under backward induction Player 2 will choose the bottom outcome on the left branch and on the right branch. On the left, this means a payoff of $12 rather than $7. Again with respect to the left branch, Player 1 will pass to Player 2, assured of getting $15 rather than $7 if choosing to quit at that node. At the preceding node, Player 2 will pass to Player 1, anticipating $12, rather than taking $6. A similar chain of reasoning applies to the right branch. Knowing the final outcomes for either the left or right branch, Player 1 then merely needs to make a choice between the left or right branch. The left branch yields a higher outcome for Player 1 and the final payoff is not only the equilibrium for the subgame, but is the subgame perfect equilibrium.

<Figure 1 About Here>

All of the discussion so far is standard for reasoning under backward induction. Now suppose Player 1 and 2 face Game B given in Figure 1. Again under backward induction, there each subgame has an equilibrium. For the left branch, it is at the third node, with Player 1 obtaining $9 and Player 2 $24. The third node is also the equilibrium for the right branch with both players obtaining $14. Given these outcomes, the subgame perfect equilibrium is located on the right branch, with Player 1 quitting at the third node.

This second game, however, has another interesting feature. Unlike the first game, depicted in Figure 1, in the second game both players could be left better off if Player 1 chose left and Player 2 quit at the second node. Of course, this would require considerable "trust" on the part of Player 1, who well understands Player 2's temptation to push the move back to Player 1. A move by Player 2 to quit at the second node would constitute reciprocated trust or trustworthiness.

The first game sketched above requires that actors know nothing about one another. Even if they understood one another's intentions, there would be no advantage to doing anything except playing so as to select the subgame perfect equilibrium. For the second game, if trusting and trustworthy behavior is possible, both actors do better than at the subgame perfect equilibrium. But in order to trust, Player 1 has to make some inference about the trustworthiness of Player 2. This goes to the heart of reading intentions. Most game theoretic approaches to understanding trustworthy behavior have started by thinking about how reputations are formed and whether a reputational signal is credible.

Reputation.

When modeling repeated games, the role of reputation becomes more obvious. (See Kreps et al. 1982, and Fudenberg and Maskin, 1986; Fudenberg and Levine, 1992). Such models assume that an actor has an initial set of expectations about the strategy that a counterpart is likely to use. The play of the game, whether repeated or sequential, enables actors to update their priors and establish credible reputations. In repeated games, the ability to develop a reputation can be critical for reaching a Pareto-superior equilibrium.

Kreps et al. (1982) point to the possibility that agents in a repeated-play prisoners dilemma game can reach one of many cooperating equilibria. Such equilibria require that at least one agent in the population have a reputation as an "irrational" player -- one who does not play the single period Nash equilibrium of defect. Fudenberg and Maskin (1986) generalize the point. In an important extension Fudenberg and Levine (1992) argue that an individual, playing for the long run, might want to invest in a reputation, even if that investment is initially costly (see also Celentani, 1996). They ask whether such an agent would do the same, even if costly actions were only imperfectly observed. Indeed they find that even imperfect observations provide some information about the long-run player's type, and from that the long-run player has an incentive to invest in building a reputation, while short-run players have good reasons for drawing inferences about that reputation. As Cripps et al. (1996) show, reputation can dissolve quickly if one or another actor holds incorrect beliefs about off-equilibrium-path play. However, there is substantial agreement that actors have an incentive to build reputations, regardless of the type of actors they face. Ordinarily these reputations involve some commitment to a form of behavior that is either perfectly or imperfectly observed (although Kim, 1996 points to ways in which "cheap talk" serves to build reputation).

While building and maintaining a reputation can be important, in a game theoretic sense this requires either a credible signal (ordinarily a player announcing a particular action over which she would not change her mind) or some observation of past behavior by the player. What remains unexplored is the source of an *initial* reputation. Any first move taken by a player when confronting a stranger requires some assessment about that player's likely actions. For single-shot games and the initial play of repeated games, reputational priors would seem to be critical. For games with multiple equilibria, the equilibrium that is reached is often characterized as path-dependent, and the initial expectations formed by partners prior to playing the game are important for defining which path is taken.

More importantly suppose there are possibilities for gains to both actors *if there is trusting and reciprocal behavior*. How can trust and trustworthiness be communicated? We argue that it has to do in part with signaling one's intention prior to the formation of any reputation.

Theory of Mind.

The idea of directly reading another's intentions is appealing, but of course is not possible. However, psychologists suggest that individuals generate a raft of signals -- ranging from the timbre of one's voice to the stance of one's body -- which betray one's intention, even if that intention is being masked. In part the argument is that humans have an evolved capacity to read the intentions of others, along with an evolved physiological set of signaling mechanisms. Research in psychology on the communication of intentions has focused primarily on autistic persons and primates (see Baron-Cohen, 1995, and O'Connell, 1998). Psychologists are interested in the capacity of individuals to put themselves in the place of others, referred to as having a "theory of mind" (TOM) about another person. This capability requires that a person be able to separate what he knows or understands from what another might.

A simplified test of this concept is often given to children. A child is shown a familiar brand of candy and is asked to guess its contents. The child typically guesses that candy is in the box, and is then shown that the box contains pencils. After expressing surprise, the child is told that another person will come into the room and will be asked what is in the box. The child is asked to predict what that person will say. Most children over the age of three believe that the response will be "candy"; they are able to place themselves into another's shoes, imagining what it is like to be the other. By contrast, autistic children invariably predict that others will think the box contains pencils. They are unable to disentangle what they know from what the other might know, or to look beyond their own knowledge and understand the mental state of the other. A "theory of mind" is a necessary precondition for predicting what others will do. An autistic person's inability to pick up on the intentions of others leads to serious problems in social interaction. The problem is not one of social ineptness, but rather a failure of the cognitive ability to imagine the mental state of another person.

There are three components to TOM (Baron-Cohen, 1995, Chapter 4). The first is characterized as an Intentionality Detector (ID). Psychologically this amounts to the ability of individuals to impute purpose and cause to another’s actions. In part this means an ability to recognize that others have goals and to derive hypotheses about how actions will be related to attaining those goals. Children as young as seven months old can distinguish between events that have a clear cause, such as a hand picking up a doll, and events with no apparent cause, such as a doll being lifted by an invisible wire (see the discussion by O'Connell, 1998, pp. 41-42). The second component involves an Eye-Direction Detector (EDD) that enables an individual to recognize the other and to draw inferences about the intentions of that other. As O’Connell puts it there are good reasons for eye detection. “The evolutionary reason why you should take very good care to detect eye gaze is because when another animal is looking at you it can mean one of the three ‘F’s. Either that animal wants to fight you, feed on you, or mate with you.” (O’Connell, 1998, p. 47). The final component involves a Shared Attention Detector (SAD). Again, the eyes are important for what is communicated. In its most mentally complex form, SAD can be characterized as a relationship between two individuals and an object. If there is a bottle of Chardonnay on a table and Rick follows Cathy’s gaze to the bottle and Cathy returns her gaze to Rick, who in turn looks between Cathy and the wine, then a complex representation can be inferred: Cathy sees that Rick see the wine (and likewise, Rick sees that Cathy sees the wine). Her request is communicated, and Rick pours her a glass. Such a complex mental concept can only occur with joint mental attention. This in turn provides an interesting guarantee of the “common knowledge” assumption crucial for game theory.

Baron-Cohen argues that a general “Theory-of-Mind Mechanism” ties together these volitional, perceptual and epistemic mental-state concepts. In turn such a mechanism enables people to draw inferences about the intentions of others by going outside themselves. Baron-Cohen 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 autistic persons appear to lack gaze-monitoring capabilities and so neither 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 short, there is a wide array of work in psychology suggesting that normal functioning humans have a capacity for reading the intentions of others. This capacity is part of ordinary childhood development. Autistic children do not demonstrate this capacity and as a consequence find it impossible to read beyond their own intentions. This implies that the structure of the brain is crucial for enabling individuals to draw inferences about the actions and intentions of others and that this capacity is not merely a process of socialization. As such, reading intentions should not be dismissed and should be incorporated into behavioral modeling of social interaction.

Emotion.

It might be argued that the rational agents in standard game-theoretic models are autistic: that is, they only require that an actor assume the other is seeking the same advantage as himself. While some models of reputation-formation and one-sided signaling offer greater complexity, very few models raise the possibility that actors read the intentions of their counterparts (and perhaps signal something about themselves) in order to coordinate on a Pareto-superior outcome.

Frank (1988) argues that emotions are useful for signaling intentions, and a reputation for having an emotional nature can yield significant advantages. For example, a reputation for vengefulness enhances a bargaining position, but a credible commitment to revenge requires a powerful emotional display. The display of anger, which cannot easily be simulated, makes credible the threat of (irrational) revenge. Emotions act as commitment devices – as incentives to behave (or not behave) in a particular manner. Guilt acts as a kind of tax on cheating, for example. Furthermore, he argues, people form judgements about others' emotional state on the basis of observable cues, such as “the pitch and timbre of the voice, perspiration, facial muscle tone and expression, and movement of the eyes….Some people we sense we can trust, but of others we remain forever wary."(p. 8) (Elster, 1999, promises to revisit much of the same territory.)

Facial expressions, among other factors, provide "telltale cues" to the underlying emotional state of a person – or of an animal. This is especially true for expressions that require muscles that are not (or not typically) under conscious control. These "reliable" cues can be used to predict cooperation or reciprocity (Chapter 6, 7). Frank presents experimental evidence that subjects accurately predict cooperation by others. (p. 137-143).

In addition to reading the emotions of others, an actor's own emotions play a significant role in much economic interaction. In a later work, Frank (1997) wonders why people would ever tip in a restaurant in a city in which they are only visiting (as well as other, seemingly non-rational behaviors). Citing Smith's *Theory of Moral Sentiments*, he contends that :

> "In this scheme, *sympathy* plays as important a role as it did in Adam Smith's scheme. I am inclined to leave the waiter in an out-of-town restaurant a tip because I feel sympathetic to the waiter's interest. I imagine myself in the waiter's position, having worked hard to provide good service, and how distressed I would feel if somebody failed to tip me. Sympathy is one of the key emotions for supporting the kind of behavior I have talked about in the examples." (Frank, 1997, p. 290).

Central to Frank's argument is the idea of the understanding by one agent of the mental state of another; this is missing in standard models of game theory. This ability to read another's mental state and respond to it may provide the basis for reciprocal behavior observed in several experimental environments [Berg, et al. (1995),Falk and Fischbacher (1998), Fehr and Schmidt (1997), Güth (1995a), McCabe, Rassenti and Smith (1998)]. In instances where there are gains to trusting and reiciprocal behavior, the ability to read others can capture these gains by facilitating and coordinating out-of-equilibrium play.

The neurological evidence strongly supports the idea that emotions are critical for internal decision making. Damasio (1996, pp. 192-194) relates the story of one of his patients with ventromedial prefrontal brain damage who drove through a raging blizzard to make his appointment. The patient recounted his skill in getting to the office and detailed the mistakes other drivers made when sliding off the road. However, when this same, hyper-rational individual was asked to decide between one of two dates for a future appointment, his decision-making apparatus "locked-up." The patient simply could not decide between the two dates, citing the benefits and costs for each date, and failing to make a choice. While very rational, Damasio notes that the patient was unable to recognize emotions and that this prevented him from making a relatively simple choice.

The amygdala, a small almond shaped organ located in the medial wall of the brain's temporal lobes and is an important regulator of emotion. A large number of experiments with primates illustrates that when the amygdala is damaged, it greatly reduces emotional controls. This carries with it important consequences for social behavior. Kling and Brothers (1992) survey the effect of damage to the amygdala on primates and a host of other animals.[2] In a number of laboratory experiments in which the amygdala has been ablated, the usual response is a loss in social rank, a cessation of social interaction and a marked decrease in "flight" behavior when confronted by another. Depending on the species, there is an expression of hypersexuality. Concerned that the laboratory environment was too alien and structured, Kling also conducted experiments on free-ranging monkeys in their natural habitat. He and his co-authors found that once the amygdala was destroyed and the animals were returned to their troop, they were unable to re-integrate. They became "social isolates" and fell prey within several weeks. This work provides evidence that, at least in animals, emotions provide an important mechanism for quickly assessing those who present a threat and those who do not. Joseph Le Doux (1996) buttresses this point: his own research shows that perceptions of danger occur very rapidly, far more quickly than conscious, cognitive perceptions of a threat can be processed.

Based on this research, it seems reasonable to conclude that, when given the opportunity, people will assess their partners in an exchange situation, and choose to trust or not based on all the information contained in the conscious and unconscious signals of others.

Social Signals.

To this point we have argued that the ability to read the intentions of others is important for humans. We have also noted that emotion may be an important source of intention -- a source that can be easily read and used to modify one's strategic action. These signals may be an uncontrolled physiological marker of an individual's affective state. At the same time there are a large number culturally derived markers that act as signals because they are correlated with patterns of play. In our interactions with strangers we often form quick impressions based on the sex of our counterpart, their status, their ethnicity or even their eccentricity. The social meaning read into these observable signals may vary considerably across cultures and over time. Even so, they systematically affect the subsequent interactions.

---

[2] Also see the survey by Aggleton (1992) on amygdala lesions in humans.

Evidence from experimental economics points to the importance of these “extra-theory” traits on strategic interactions. Actors may treat a subject's "type" as a signal of their expected play -- in effect, a signal of intentions. Several studies investigate behavioral differences between men and women in the sorts of games described above. The argument common to these studies, drawn from theories and evidence from social psychology, is that women are expected to be more "cooperative" than men. In the ultimatum game, for example, women are expected to reject an offer of a given size less often than men. If players anticipate this behavior, then both men and women would offer less to women than to men. Eckel and Grossman (1998a, 1998b) find some evidence of both conjectures -- women do reject less frequently, and all players offer (slightly) more to men than to women. (Eckel and Grossman (in press) survey results on ultimatum, dictator, and public goods games.)

Another factor that might be expected to affect strategic play is social status. Ball, et al. (1998) and Ball and Eckel (1998) examine the effect of artificially-induced, observable status differences on play in market and bargaining settings. (Their procedure awards gold stars to a subset of the players, who then wear the stars while playing a subsequent game.) In general they find that subjects who are awarded higher status have higher earnings in the games. Eckel and Grossman (1996) investigate another sort of status difference. They examine the effect of a credible signal of neediness, and find that subjects are more generous to needy subjects in a dictator game setting. Status differences appear to affect the expectations of both high-status and low-status players. Players make inferences about the intentions of their counterparts based on the status differential.[3]

Signals embodied in traits are beyond the control of the signaling actor, but many other social signals are (largely) deliberate or at least contingent on the situation. A social signal that is quite obvious (and may or may not be strictly controlled by the expressor) is the human face. 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 displays) are evolved and serve a function for the species. So a peacock's ostentatious display of his tail or a chimpanzee’s baring of her teeth is 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

---

[3] In addition to the experimental work in economics, there is a very large literature in sociology on status characteristics and influence. See, for example, Ridgeway (***), Webster (***), ??

occasion other species as well. We now turn to a discussion of faces as social signals and argue that they provide one important way of thinking about signaling (and reading) intentionality.

Facial Expressions

An entire industry has grown up around the study of facial expression. Much of the research has followed the lead of Ekman (1972; 1983) and focuses on Darwin's first argument that facial expressions are evolved. If expressions are evolved, then humans must share a common, universal set. Many facial expressions are involuntary in non-human animals, and probably in humans as well. As a consequence, much of the research has turned toward understanding what facial expressions reveal about the underlying emotional state of the expressor. If facial expressions constitute emotional leakage, then the emotional content should be obvious to others because everyone shares the same universal repertoire of expressions. Of course, learned social behavior may work to mask those emotions; because social experience varies across cultures, cultural differences may lead to different forms of masking.

There is some support for this position. Researchers have found that subjects label certain expressions of emotion in the same way regardless of culture (see reviews by Ekman, 1983; Fridlund, Ekman and Oster, 1987). By and large this research has uncovered possibly six distinct emotions that are read with ease across cultures. These emotions include anger, disgust, happiness, sadness, fear, and surprise. As Fridlund, et al., note, at a minimum the emotions of happiness, anger, disgust, sadness and combined fear/surprise have been validated in both literate and pre-literate societies (1987, p. 159). Typically subjects are shown photographs of posed expressions, with the poser from a different culture (and often from a different ethnic group). Subjects are then asked to match an emotional expression with each photograph. These results have been shown to be fairly robust (although see Russell, 1994 and the discussion below).

The brain reacts differently to different facial expressions. Damage to specific areas of the brain limits the capacity of humans to recognize emotional content. Morris et al. (1996), using positron-emission tomography (PET), find that the neuronal response in the left amygdala is greater when subjects are given an image with a "fearful" expression than with a "happy" expression. Moreover, this neuronal response significantly interacts with the emotional intensity of the image. Morris et al. conclude that this is strong evidence that the human amygdala is actively engaged in processing the emotional salience of faces. Breiter et al. (1996) make the same point, imposing additional controls over the facial stimuli and noting that the human amygdala responds preferentially to high valence emotions like happiness and fear. Similar findings have been found among primates whose amygdala have been surgically ablated (Weiskrantz, 1956; Aggelton and Passingham, 1981; for humans, see Adolphs et al., 1994).

In a case study by Young et al. (1996), a woman with a partial bilateral amygdalotomy showed considerable impairment for recognizing facial expressions. This was true to the extent that she often mistook the same individual with different emotional expressions as different people altogether. Adolphs et al. (1996) asked a sample of patients with focal brain damage to assess emotional facial expressions. They used brain imaging techniques to map "successes and failures" in gauging emotional content, finding that all of their subjects recognized happiness, but that subjects with lesions in the right hemisphere were impaired in identifying negative emotions -- particularly fear and sadness. These findings indicate an important functional connection between the way the brain is wired and whether expressions are recognized. However, they do not say much about the emotional state of the subject.

Generally, two positions are taken with respect to facial expressions. On the one hand, facial expressions signal a particular emotional state held by the expressor. This perspective leads to the conclusion that there are identifiable states that are easily perceived. On the other hand, the facial feedback hypothesis contends that facial expressions activate affective states. If I smile, this elevates my emotional state, making me happy rather than angry or sad (see the discussion by Izard, 1997).

Criticisms abound. Russell (1993, 1994) points out that the universality of expressions may not be quite so universal. Moreover, the forced choice mechanisms used in assessing facial expressions may be the cause of much of the cross-cultural agreement about the content of an expression. To get some sense of this problem, we briefly look at the most readily identifiable facial expression -- the smile -- and judge some of the evidence for what the smile signals.

Smiles.

The smile represents the emotion that is the most easily recognized: happiness. Typically this expression is given by a symmetric contraction of the *zygomatic major*, slight exposure of the teeth and wrinkling of the *orbicularis oculi*. The usual finding is that more than 80 percent of different populations agree that smiling photographs signal happiness. However, while a smile generally yields a clear emotional meaning, this does not imply that all smiles are equal. A number of experiments suggest that “false” smiles may be difficult to judge from still photographs.

In an early paper discussing smiles, Ekman and Friesen (1982) argue that false smiles usually do not involve the "crinkly-eye" effect where muscles at the corner of the eyes contract, such smiles tend to be asymmetrical, they occur at inappropriate times and they have excessively long duration.

Much of the subsequent literature has had a difficult time uncovering differences between felt and false smiles. Ekman et al., (1988) designed an experiment to test whether such expressions can be delineated. They asked a group of student nurses to observe segments of videotape and then deliberately mask their emotional feelings when speaking to an interviewer. At the same time, the faces of the subjects were videotaped for later analysis. One stimulus presented the subjects with a nature film that was peaceful and serene, while the other stimulus showed amputations and burns and was designed to elicit strong unpleasant emotions (p. 415). Using their system of Facial Action Coding, coders checked types of smiling expressions. In their analysis, Ekman et al. find there were subtle differences among forms of smiling depending on the condition to which subjects were exposed -- these had to do with expressions with muscular activity around the eyes. If only smiles in general were taken into account, then there was no difference between truthful and deceptive behavior. In a subsequent study, which again presented film clips, but this time secretly recorded subjects and left them by themselves, Ekman et al. (1990) find predictable differences in the type of smile -- "felt" smiles versus "masked" smiles. At the same time subjects were measured for brain electrical activity. While some differences were found between felt and masked smiles, there was no significant difference between the felt smiles and the baseline condition. Nonetheless, the interesting finding is the extent to which subjects engaged in smiling activity independent of others present in the room.

Returning to the videotapes of nurses, Ekman and O'Sullivan (1991) culled 10 different segments involving different subjects. Half were "lying" in response to the interviewer, while the other half were telling the truth. These clips were shown to different audiences and those individuals were asked to judge whether the speaker was "honest" or "deceptive." For two of the ten clips, these individuals were also asked to briefly describe why they thought the person on the videotape was either truthful or lying. The audiences ranged from members of the Secret Service to Federal polygraphers to college students. Interestingly they find that members of the Secret Service are best able to differentiate between the honest and deceptive video clips. More generally, those who are best able to discriminate between the two report that they rely heavily on nonverbal behavior as well as subtle facial expressions.

This led Frank and Ekman (1997) to look more systematically at the detection of deception and to tie this more explicitly to the recognition of facial expressions. In the general experimental design, subjects are assigned to conditions in which they (or a confederate) have an opportunity to take $50 from a briefcase. Each subject is then interrogated (from scripted questions) by a different subject about whether he took the money from the brief case. All subjects in the "stealing" condition are told to deny they have taken any money -- whether they did so or not. If the interrogator guesses they are lying, then subjects are penalized whether guilty or not. Punishment involves losing the $50 (if it was taken), losing the show-up fee of $10, and being subjected to randomly timed blasts of white noise in a darkened room (this latter was never carried out).

In addition to the high-stakes condition described above, a low stakes condition has the same subject tell the truth or lie about his opinion on several issues to the same interrogator. The experimenters of course know which subjects are lying. Punishment is minimal in this setting. Videotapes of the subjects are taken and coded. Ten different instances of the high stakes (theft) and low stakes (opinion) treatments are presented to a new group of subjects who are asked to judge deception.

In their first experiment Frank and Ekman find that observers' accuracy in judging deceptions is correlated across the high and low stakes setting. This means there is variation in the capacity of individuals to assess deception. A second experiment replicates the first and then adds a second treatment in which the new observers are asked to judge the emotional content of faces flashed at 1/25th of a second. Here the authors find that there is a positive correlation between the ability to detect deceit and the ability to identify "micromomentary" facial expressions of emotion. In other words, those subjects best able to detect cheaters also are best able to read faces.

Carroll and Russell (1997) take the point further. In their experiments they find that subjects do quite well when they can judge a facial expression within a particular context. In the Frank and Ekman study noted above, when subjects were able to observe the way in which smiles or other expressions unfolded, they were much better able to judge the emotional content of the expression. Carroll and Russell go on to argue that reading emotions is highly associated with a number of component processes, including the social context and the cognitive capacity of the expressor. As such, "All facial movements are the direct outcomes of these component processes. An emotion is therefore expressed in the face only indirectly, through its correlation with the components." (p. 165). They make a strong claim for facial expressions as social signals that can be intended.

Fernández-Dols and Ruiz-Belda (1995, 1997) make this same point, and in their experiments they find little support connecting happiness and smiles. Their main finding is that happiness "*is not a sufficient cause of smiling*" (italics in the original, 1997, p. 269). Instead smiling is a joint effect between happiness and social interaction. As such, smiles become important means for communicating a signal to others and are expressed more often in clear social situations. Their 1995 study of Olympic Gold Medalists analyzes the expressions of winners coming to the podium to receive their medals. Within a five minute period there was a marked change in facial expressions -- even though those Gold Medalists were presumably very happy -- with smiles recorded the most during a period of social interaction and the least during periods of individual isolation.

Two positions are taken with respect to truthful and false smiles. One holds that smiles are, again, a reflection of emotional leakage, and that truthful smiles can be discerned. The evidence supports the point that subjects are good at judging the spontaneity of a smile within the context of that expression. From the standpoint of Ekman and others this means that there are micro-expressions that occur just prior and subsequent to a smile that enable subjects to discern whether the smile is genuine. The other position holds that it is the context of the smile that gives away its meaning (and this is the point made by Fernández-Dols and Ruiz-Belda). In this sense a smile is a form of social signaling and it can be discerned only by accounting for the audience and the environment within which it is given. Both positions cast doubt on the value of still photographs for eliciting consistent assessments of the emotional content of a smile. This point will become important in the discussion of our results below.

Facial Social Signals.

The discussion above suggests that facial expressions unfold in a context. Both the expression and the context constitute important signals for an individual trying to gauge something about the intention of the expressor. This seems to indicate that facial expressions must do more than leak emotion. They must be part of an important social signaling mechanism.

Evolutionary psychologists propose that facial expressions, which are largely linked to the physiology of our musculature, are traits that have been evolutionarily selected. In turn, mental modules that recognize those expressions and activate particular behavioral responses also have an evolutionary basis. One such stylized story told by Fridlund (1994) supposes that early hominids sought to minimize expenditure of effort when confronting one another. Meetings between individuals from different groups would typically precipitate a fight, which in turn is a costly activity for both the winner and loser. When vocalizing a warning, facial muscles contract in a way resembling what we would consider anger. The more exaggerated that facial expression, the better able another is to read it. Interlopers who develop a "friend or foe" detector are more likely to flee. In such an instance both parties are likely to preserve their strength, eschewing a fight in which one or both are injured. A similar story could be told for the "friendly" aspect of this particular mental module.

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 little investigation of behavioral responses (however, see Knutson, 1996). The consequences of facial expressions for behavior in different social settings are an open question.

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 intentions.

While the study of social signals has not dominated the study of facial expressions in adults, it has been a mainstay in the child development literature. From the outset, faces appear to be crucial for child development. In an imaginative study Johnson et al. (1991) trace the reaction of newborn 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 uses a plexiglass table, half of which is transparent, creating a “visual cliff” in the center. The infants are placed on the solid surface and at the clear end of the table stands the infant's mother. The majority of the infants crawl across the clear space when given a posed smile by the mother, whereas they do not if the mother produces a posed fearful expression. Obviously facial expressions evoke important social signals that are well understood at a very young age.

Abstract Images.

While we emphasize that human facial expressions serve to signal intentions, it should also be clear from our discussion that snapshots of facial expressions are difficult to "read." That is, the emotional content of an expression is often unclear or can easily be misinterpreted (or even misrepresented).

Human facial expressions can be very complex; the muscle groups on the face can send a wide spectrum of signals. While researchers have focused on the six primary emotions -- happiness, sadness, anger, fear, surprise and disgust -- humans are capable of forming 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 susceptible to systematic evaluation. Taking this insight, McKelvie (1973) designed an experiment in which he used schematic representations of faces. These schematics resemble variations on the ubiquitous "happy face" wishing everyone a nice day. McKelvie uses an oval to represent a head and then draws in line segment representations of eyebrows, eyes, nose and mouth. These are systematically varied and then presented as stimuli to subjects.

A total of 128 schematic faces are used; each subject is presented with a sample of 16 faces. Working one at a time, subjects are asked to rate how easy it is 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 reflect 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 expressions signal 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) use schematic drawings of facial expressions thought to represent the six primary emotions, as well as selected photographs from Ekman and Friesen's "Pictures of Facial Affect" (1978). In one of the experimental conditions children are 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 is significantly greater for the schematic drawings than for the photographs. Accuracy varies, however, with children having the easiest time identifying happiness, sadness and anger (p. 383). The simplifications of the schematics are readily apparent to these children and the emotion evoked is 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. In a subsequent section of this paper we will rely on these results in order to concentrate on relatively clean expressions.

Survey and Experiments with Human Faces

We are interested in whether facial expressions signal intentions about either trusting or trustworthy behavior. We argued above that facial expressions should produce social signals and we investigate these within the context of a game theoretic task. Our first experiment focuses on human faces.

In the first experiment we did two things. First, we photographed two models, a young male and a young female, in several posed facial expressions. These expressions were then presented to two large undergraduate economics courses. Subjects in these classes evaluated the content of the expressions. Second, subjects in five large undergraduate classes were presented one of these expressions and asked to make the first move in a two-branch trust game. Consequently, we have evaluations of the images from a large population and we have a large pool of behavioral responses to the play of a trust game.

Survey of the Human Facial Expressions.

We began by creating our own facial expressions. Two summer interns with the National Science Foundation were selected to pose. Given our experimental design we thought that choosing college-age models would enhance our credibility in the experiment described below. The two models were told to pose three distinct facial expressions: happiness, neutrality and anger. From a large number of digital photographs, the authors selected representative pictures selected and cropped them to the same size.

A total of 154 subjects in two large introductory economic classes at Virginia Tech were used to judge the emotional content of those pictures. Six photographs were displayed, each representing a specific emotional expression for the two models.[4] The image was presented to the subjects for 15 seconds; subjects were then asked a series of questions about that image. The questions were posed as statements (e.g., the person in the picture is happy) and subjects were asked to use a 5-point Likert scale in responding to the statement. Responses ranged from (1) Strongly Agree to (5) Strongly Disagree. The items ranged across the usual set of emotional affect items; three are targeted here: happy, sad and angry. Subjects in the first group were shown the male pictures first, running from the smiling to the neutral to the angry expressions. These images were followed by the female expressions in the same order. For the second group we began

---

[4] Although we intended to test all three posed emotional expressions -- happiness, neutrality and anger -- in the game theoretic experiment we were unable to conduct a condition with the female angry face. To keep a balanced design we omit the discussion of the male angry face

with the neutral expression, then the angry and finally the smiling expression, again with the male presented first and followed by the female image. A third session was scheduled which would have begun with the angry expression, but we were unable to run that session.

Figure 2 presents the average ratings of emotional content for the pairings of smiling and neutral images. The obvious finding is that subjects are quick to pick out happiness as the emotion portrayed by the smiling models. Moreover, there is little difference between the male and female image.

<Figure 2 About Here>

This point is borne out by an ANOVA in which happiness is estimated as a function of the sex of the image, whether or not the image expressed a smile and the interaction between the two. The main effects were significant (for sex, F=16.04, df=1, p<.001 and for smile, F=8588.14, df=1, p<.001). Moreover, the interaction between sex and smile was significant (F=2628.78, df=1, p<.001). If subjects viewed a smile, they rated the model's emotional state as happy. If the model was female, subjects rated her as more happy, and if the model was a female and smiling, she was more likely to be rated as happy.

Likewise if a similar ANOVA is run for "angry" then there are significant main effects for sex (F=7.69, df=1, p< .006) and smile (F=765.55, df=1, p<.001). At the same time the interaction is also significant (F=241.73, df=1, p<.001). Interestingly, in post hoc tests, subjects are more likely to find the female angrier than the male under the neutral expression.

Finally, if we compare the difference between sad and angry, we find that only the sex main effect is significant at the .05 level (F=7.24, df=1, p<.007). Neither the presence of a smile (F=3.52, df=1, p<.061) nor the interaction between sex and smile (F=.63, df=1, p< .429) is significant.

From these results it appears that our facial images clearly differentiate between happiness and other emotions. Subjects are quite adept at reading something about the emotional state of the other, although the female is generally regarded as happier when smiling and more angry when neutral than her male counterpart.

The Game Theoretic Experiment.

In the experiment, five additional large introductory economics classes were used. Four of the classes were shown a specific facial expression from our two models. In the remaining class, our control group, no image was displayed.

The design partially replicates that found in McCabe, Rassenti and Smith (1998 -- hereafter MRS). In their experiment a group of subjects plays the extensive form game found in Figure 3. The problem posed here is identical to that of Game B in Figure 1. Each box represents a decision node for a particular player, whose number is in the box. Payoffs for each terminal node are indicated in the ovals, with Player 1's payoff appearing first. In the experiment payoffs were given at the rate of $.10 (US) per unit.

<Figure 3 About Here>

As MRS argue, this game has special features that asks Player 1 to make a trusting move. The initial choice for Player 1 is between the left and the right branch. That choice is based on beliefs about Player 2's intentions. A choice of the right branch leads to an outcome of (40, 40) and requires only self-interested play from Player 2. A choice of the left branch can lead to a superior outcome for both at (50, 50), but only if Player 2 reciprocates.

This game has the virtue of containing an initial move in which Player 1 can enter either a competitive game (the right branch) or a trust/reciprocity game with a cooperative payoff (the left branch). Our conjecture is that a smiling facial expression signals an intention to reciprocate trusting behavior and as a consequence subjects should choose the left branch more often when observing a smiling counterpart.

Our experiments replicate the instructions and general layout of the MRS experiment, although with several differences. First an entire classroom of subjects was given the instructions by the same experimenter with the instructions projected on a screen at the front of the classroom. The script and instructions informed the subjects that they were participating in a pre-test of an experimental design. Subjects were told they were being paired with another subject who had already played the game, and whose choices had already been recorded. They were told that five participants would be chosen at random and paid for their decisions.

At the conclusion of the instructions, depending on the treatment variable, subjects were shown an image of their counterpart and a sheet was handed out with the decision problem. That sheet resembled Figure 3, and included boxes to check for alternative choices down the decision tree. All subjects were assigned the position of the first decision maker and the first choice was whether to move left or right. Once subjects made the first choice the experiment was halted, and five of the subjects were randomly selected to complete the experiment and receive payment. Play was completed privately outside the classroom. The "counterpart's" play in subsequent moves of the game was based on the distribution of play for a single-shot game as reported by MRS. The rest of the subjects completed a brief questionnaire and manipulation check. After all materials were collected, subjects were debriefed on the design and purpose of the experiment.

<Figure 3 About Here>

The primary manipulation involved providing a fixed cue to each subject in the form of a facial image. Subjects were shown one of five cues previously given in Figure 2. Four groups were presented one of the images shown to the class room groups which assessed the emotional content of the images. A fifth group was used in which no image was shown to the subjects. This control group is directly comparable to the MRS one-shot experiments.

Analysis

Our principal conjecture is that "friendly" social cues are critically important for initiating a cooperative move. Consequently we expect faces that are readily interpreted as “happy” will increase the likelihood that subjects take a trusting move at the first node of the game.

This conjecture is only partially supported. Table 1 presents the percentage of subjects taking a left (trusting) move in the game. To begin, our results are different from those reported by MRS. The first column shows the percentage of subjects choosing a trusting move given that they observed no facial image. This is the control condition and is the most similar to the MRS design. With no image, only 39.5 percent of the subjects chose the left branch. By comparison, when the male smile image was used, there were increases in the percentage choosing a trusting move.[5] The same was true for the neutral female image. There is a marked decrease in the percentage choosing to trust when exposed to a neutral male image or the smiling female image.

<Table 1 About Here>

Also presented on the table are the percentages of subjects assessing the image as either friendly or cooperative. When subjects completed their initial decision -- whether to choose left or right -- they were asked a series of questions about their counterpart in the experiment. The questions offered a statement (“My counterpart (in this experiment) was:”) and then gave subjects three choices. For the “friendly” item, they could choose friendly, neutral or unfriendly; for the “cooperative” item, they could choose cooperative,

[5] As a manipulation check subjects were asked afterwards whether the image they viewed was that of a male or a female. A total of 29 subjects incorrectly marked their form -- indicating the image was different from what they observed. These data were not included in the analysis discussed here because these subjects did not meet the minimum criteria for acknowledging that they had observed the stimulus. Other analysis indicates that including or excluding these subjects makes no substantive difference in the findings.

neutral or unfriendly. By and large these ratings are consistent with the images, but do not help explain the fact that the results for the female images are opposite what we expected.

Table 2 presents a multivariate analysis of the choice to go left or right. Subject's choices are estimated using probit and we examine three models. The first model tests whether there is a difference on average between the control group and the groups with facial images. There is no difference, but this much is clear from an examination of Table 1. The second model looks at the main effects of the male/female model and the smile/neutral facial expression. The intercept term constitutes the blank control group, and the negative sign indicates that on average, subjects were less likely to choose the left (trusting) branch. When viewing the female face, subjects were somewhat more likely to choose the right branch, but this main effect is weak. The same is true with the weak effect of the smile. Once we introduce the interaction term of the female*smile, both of the main effects and the interaction effect are strong. Again, this is consistent with what is observed in Table 1 in which the female image elicits greater trust, the smile elicits greater trust, but the smiling female results in much less initial trusting behavior. Even if assessments of the image are included in estimates, these findings are unchanged,

<Table 2 About Here>

It appears from these data that smiling expressions affect decision-making. However, subjects appear to be less trusting of a female smile, even though they rate the image as cooperative. Perhaps subjects do not believe that the female smile is a true smile. It seems clear from these data that a mixed message is being sent from the still photographs. This finding is consistent with other work in the literature noting the problem with assessing false smiles.

Survey and Experiment with Schematics

The facial expressions used in the first experiment were posed, not carefully controlled, and were limited to two different models, one male and one female. While we found that faces make a difference when considering a trusting move, the extent of these differences is not clear. There are several concerns with the first experiment. First, subjects may not have felt they had a real stake in the problem. Only a few subjects were selected from each group to be paid and their stakes were relatively small. Thus, they may not have taken the task seriously. Second, they may not have believed they were participating with the individual whose facial image was shown. If so this, too, could imply that they did not take the task seriously. Finally, even though they were taken through a lengthy set of instructions, they may not have understood the complex decision problem.[6]

To correct for this we designed a second experiment that relies on an abstraction of different facial expressions. In this setting we strip away the complex array of signals that can be formed by the human face. Instead we focus on a limited set of expressions and at the same time turn to a set of items that tries to tap the trustworthiness of the images. In addition, we simplify and streamline the instructions.[7] To get at what our stimuli mean, we survey the same subject pool as in our previous experiment -- Virginia Tech undergraduates. We then turn to a controlled laboratory experiment in which subjects are assigned a specific schematic representation and then engage in a trust/reciprocity game with another subject.

The Survey.

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. These classes consist primarily of college sophomores; about 1/3 are 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 nine icons and asked to rate its characteristics. The icons are based on a 3x3 design involving three manipulations of the mouth and three manipulations of the eyebrows, as shown in Figure 4.

<Figure 4 About Here>

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 was randomly assigned for the word pairs. In the analysis presented here we focus on ten paired items from the instrument: Five items relate to a behavioral

---

[6] There is some evidence that they did not comprehend the instructions. In a question asked following the decision, 35.9 percent of the subjects report that the instructions were unclear.

[7] Under the new experimental design we asked subjects whether they thought the instructions were clear. Just under 97 percent found the instructions to be clear.

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.). Figure 5 plots the mean response across four of the icons for each word pair. Each icon's mean for the word pair is 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. From left to right, the *happy* face is judged on average to be more trusting and trustworthy than the *sad* face, the *angry* face and the *devious* face. What is clear from the figure is that the order is preserved across nearly all ten items.

<Figure 5 about here>

The rough "eyeballing" of mean results pretty much tells the story. Across the first five "behavioral" items, the differences among 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 same is true with respect to affective (emotive) items. 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 is ordinarily thought of as angry and devious affective states. Moreover, these findings are consistent with those found by other researchers who have used either schematic or facial images.

The first five items and last five items were combined to generate two scales. Both the "behavior" and the "affect" scale add an individual respondent's score across the items contained in Figure 5. An average score across these items was then calculated for each respondent. Two separate models were then estimated for each scale as shown in Table 3. Four independent variables are included as well as an additional control variable; Model 2 includes interaction terms. The variable SMILE is a dummy variable for icons with an upturned mouth. Likewise FROWN is a dummy variable for a down-turned mouth. UPBROW is a dummy variable for eyebrows that are upturned at the center (/ \)and DOWNBROW does the same for down-turned or "frowning" eyebrows (\ /). The neutral mouth and brow are then reflected in the intercept term of the regression. In model 2, these variables are interacted as indicated. Finally we add a dummy variable for the SEX of the respondent, controlling for perceptual differences that might emerge from assessing these icons.

<Table 3 About Here>

Both models confirm findings by McKelvie (1973) and others noted above. The positioning of both the mouth and eyebrows significantly affects the assessment of the facial icons. With respect to the behavioral scale (trust, honesty, etc.) the intercept term reflects the midpoint of the general semantic differential scale and is consistent with what we might expect from a neutral icon. The effect of eyebrows is pronounced. When the eyebrows are upturned, they decrease the evaluation (move it toward the "trustworthy" end of the scale) by almost a full point. Down-turned eyebrows have the opposite effect. Interestingly, the mouth position does not affect behavioral assessments except in interaction with the eyebrow positions. In Model 2, the interaction between Smile and Downbrow has a strong positive effect on the index, indicating a move toward the "untrustworthy" end of the scale. A smile does not have an unambiguous effect on the perception of the likely behavior of a facial icon. The effect can be positive or negative, depending on the position of the eyebrows. Main effects alone would lead one to believe that the frown/downbrow combination would be most negatively perceived, but interaction effects adjust the evaluations so that the conflicting message of the smile/downbrow icon is perceived as suspicious and dishonest.

When turning to the affect scale, both the smile and the eyebrow positions are strongly related to the evaluation of the icon. These results are consistent with the behavior scale in that a smile and upturned eyebrows result in a more positive assessment, while a frown and down-turned eyebrows lead to a more negative assessment. Interaction terms are again strong, with the combination of a smile and downturned brows having a significant effect on the evaluation of the icon. What also appears from the estimation is that female respondents are more likely to evaluate the icons harshly with respect to affect. The effect is not large, but is statistically strong.

In short, these estimations reinforce the results from Figure 5. The position of eyebrows and mouth both matter for the inferences that respondents draw about the icon. Put another way, respondents are drawing meaning from the icons, and the meaning is systematically related to our manipulations. Subjects differentiate between the various icons, and the icons can be ordered from most to least *positive* on both behavioral and affective characteristics. These results set the foundation for the subsequent experiment.

The Laboratory Experiment.

In this experiment pairs of subjects participated in series of two-person sequential games. All games involved sequential choice under perfect and complete information. In a typical game each subject made one or two moves. A total of 168 subjects were recruited from the local student population at Rice University. Students were contacted in their dining halls and asked to volunteer for a decision making experiment. Subjects signed up for one of twenty-three planned experimental sessions. (Fourteen sessions were conducted in February, 1998; an additional nine were added in November, 1998).

The laboratory accommodated eight subjects, each seated in a cubicle formed by moveable partitions and facing a computer. Although subjects were in the same room and could hear one another, they could not see one another's computer screen. At the outset of the experiment subjects were cautioned not to speak and told that if they did so, the experiment would be canceled. All experimental sessions were conducted over local area network, and the network handled all communication between subjects.

Upon arriving at the laboratory subjects choose their seats and were asked to sit quietly until all the volunteers arrived. At least nine subjects were recruited for each session in order to ensure that eight participated. If more showed up, a volunteer was solicited and paid $3.00 on the spot to leave. If no one volunteered, one subject was randomly selected, paid and dismissed. Once the requisite number of subjects appeared, the experiment began. In eight of the twenty-three 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 were given self-paced instructions and shown how to make choices in the sequential game.[8] These instructions are attached as Appendix 1 and differ somewhat from the instructions used by MRS.[9] In each experimental session, subjects participate in as many as 30 games.[10] Prior to each decision subjects were randomly rematched. Because of the limited number of subjects, same-pair play often occurred. However, subjects carry no unique identification in the course of the experiment, so it is impossible for subjects to know with whom they were paired in each game.

The same procedure was used in each period of play. Before each game began, subjects were told to which role they were assigned (1 or 2) and who would move first. Each subject was 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 was then

---

[8] 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.

[9] 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.

[10] 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.

revealed, depending on the icon manipulation as explained below. In the games analyzed here, player 1 always had the first move and player 2 waited until the first player's choice was complete. At the first move, Player 1 chose either the left or right branch of the game. At subsequent moves a player had two choices -- either to quit the game and take the payoff or to pass to the next player. Once a choice was made, the other player was notified of the move. If the choice was to quit, the payoff box was circled on the computer screen, and both players were notified of the outcome and asked to record their payoffs for that period. If the move was passed, the next player made a choice while the other player waited. The computer mediated the only communication between players and subjects were only told the moves of their own partners. If the game continued to the last decision node, the second player had the choice of two payoff boxes. Once the game ended, subjects were instructed to wait until all pairs completed their own decisions. At that point the subjects were again shuffled and re-paired.

Participants were paid in cash and in private at the conclusion of the experiment. The games used in this analysis are identical to those given in Figure 1. Payoffs in those games 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. The program randomly selected a period and informed the subject of the period drawn and earnings. Subjects were asked to verify that the payoffs for the draw matched what they 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 40 minutes. On average subjects earned $13.47 for their participation.

Icon manipulations. Subjects know little about their partners in each period. 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. Seven distinct manipulations of identity were used and five distinct pairings of icons were used. In each session half of the subjects were randomly assigned one of the icons in the pair and the remaining half were 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 has been assigned an icon at the outset and that each player retains his or her assigned icon for the duration of the session.

We deliberately did not tie the icon to any personal characteristics of subjects. They were simply told the icon assigned to them was 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 and without human facial characteristics. Sessions are of three types: those with facial icons, with nonfacial icons, and with no icons. Our expectation is that the icon type of a player and the partner influences trusting and trustworthy behavior. Four facial 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 is characterized as "devious;" the icon with upturned eyebrows and an upturned mouth is characterized as "happy;" the icon with downturned eyebrows and a downturned mouth is characterized as "angry;" and the icon with upturned eyebrows and a downturned mouth is characterized as "sad." Two additional icons are used that have no human facial content. Here a rectangle and an oval 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 the program selects one. 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.

We introduce a final control condition in which subjects have no information about their counterparts. Their "icons," for all intents and purposes, are blank. 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.

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 the February 1998 data in subsequent periods a game was randomly selected for each pair in each period. In the November 1998 data the same fixed order for all games was used for all pairs of subjects.

In this analysis we limit our attention to two of the 18 games.[11] These games are the same as given in Figure 1 and were selected to tackle two concerns. Game A was chosen to assess whether subjects understood basic properties of backward induction for this experimental design. As noted above, the unique subgame perfect equilibrium is the last node on the left branch.[12] In this game subjects have no incentive to use strategies that yield out-of-equilibrium behavior. Neither trust considerations nor concerns about reciprocated trust should enter into subjects' strategic calculations. There is no incentive for subjects to do anything except to play their narrow self-interest.

Game B is quite different. It is identical to the "trust" game used in the first experiment, except that the payoffs have been transformed. In this game the unique sub-game perfect equilibrium is the middle node on the right branch. As before, the trust/reciprocated trust move is the upper left node.

Analysis.

We first ask whether subjects understand the basic structure behind the game. Game A is straightforward, and the Nash equilibrium outcome should dominate. Figure 7 provides descriptive statistics for the node at which subjects ended the game. Indeed the Nash equilibrium was chosen 81.3 percent of the time. Another 12.5 percent of the time player 1 chose the right branch, and in that instance all subjects went to the equilibrium for that subgame. Only a handful (3 subjects) made anything equivalent to a mistake -- that is, an outcome inconsistent with equilibrium play. All in all, subjects appear to understand backward induction and behave according to game theoretic predictions, when it is in their interest to do so.

---

[11] The design of the experiment tested a variety of sequential bargaining games with differing payoff parameters. Eckel and Wilson (1998; 1999) analyze the bulk of these other games and explore the history effects associated with these different games.

[12] To ensure there was no selection bias, branches to this game were randomly flipped throughout the experiment.

<Figure 7 About Here>

A different picture emerges for Game B. Here the game-theoretic prediction is chosen only 32.4 percent of the time. Instead the first player takes a trusting move 64.8 percent of the time, choosing the left instead of the right branch. Moreover, that trust is generally reciprocated. Taking into account only moves onto the left branch, just over 70 percent of the time the second player chooses to quit, even though that player could pass the move back to player 1 and anticipate a larger payoff.

Does the introduction of an icon make any difference for trusting behavior in Game B? In the sessions where faces are presented to the subjects, 69.9 percent of first-movers choose to trust their counterparts; in sessions with nonfacial icons or blanks, 57.5 trust on the first move. This difference is statistically significant at the .1 level ($\chi^2(1)=2.86$, $p=.091$). When faces are present, there is more trust than when no faces are present.

Table 4 presents the raw percentages of trusting moves broken out by pairings of icon types. The row positions indicate the first mover while the column positions indicate the second mover. This table gives a rough sense of the first mover's action conditional on the pairing. Again we see the pattern observed above: there is generally more trust in sessions with facial icons than without. The lowest level of trust is seen in the sessions with no icons at all. Interestingly, Happy and Sad do the most trusting, perhaps because they expect trust to be reciprocated.

<Table 4 About Here>

To more carefully investigate the question we used a series of models estimated using probit analysis. These models are reported in Table 5. The first model estimates player 1's initial choice: whether to move left or right. This is estimated as a function of whether or not the player observed an icon with a facial characteristic and at what point in the experiment the subject encountered the game. Because subjects played a number of periods, there may be supergame effects that are a function of the number of decisions made. For this reason we include the variable Period as a proxy for supergame effects and/or learning.

<Table 5 About Here>

While the results are statistically insignificant, the patterns are consistent with expectations: the variable Face carries a positive sign in the first two models, and in combination with the Smile and Eyebrow variables has an overall positive impact in the third. Here is no apparent in-group effect, as indicated by the "paired" variables, one for nonfacial icons, and one for facial icons. Most interesting is that the interaction term (in this case the "happy" icon) has a negative effect on trusting. That is, when observing a partner who holds a happy expression, the first mover is less likely to take a trust move.

The preceding analysis concerns trusting behavior: what of reciprocated trust? The second player has the opportunity to reciprocate by choosing to quit if the first player picked the left branch. Once again it appears that the presence of an icon with a facial expression matters. In these experimental conditions 79.2 percent reciprocated the first mover's trust. By comparison only 54.8 percent of subjects in a condition with no facial expression reciprocated. These differences are statistically significant at standard levels ($\chi^2(1)=7.55$, p=.006).

Table 6 presents Probit estimates for the next move of the game: the decision to reciprocate. The analysis only includes pairings in which the first mover chose to trust (picked the left branch at the first move). The dependent variable is estimated as the probability that Player 2 reciprocates a trusting move..

<Table 6 About Here>

Not surprisingly, there is a strong effect for "Face": In sessions with facial icons, there is significantly more reciprocal behavior than in sessions with nonfacial icons or blanks. In the second model we explore whether there is reciprocated behavior which is a function of in-group/out-group effects. We find no significant in-group effect when nonfacial icons are paired (either a rectangle with a rectangle or an oval with an oval) nor when facial icons are paired. However, there is a positive, strong effect of reciprocating when there is no signal (the blank condition). In the third model, we find there are no main effects for either the smile or the eyebrows.

Finally, what happens when the first mover has trusted, but the second mover fails to reciprocate that trust? This happens with some frequency (29.8 percent of the time). More than half of the time (61.8%) the first player "punishes" the second player for failing to reciprocate. In 73.3 percent of the cases, subjects with facial icons punish their counterparts with facial icons (11 of 15). By contrast, subjects with no facial icon or with no icon whatsoever, punish only 52.6 percent of the time (10 of 19 cases). This behavior is consistent with that found by Eckel and Wilson (1998) when looking at "rejected" asymmetric offers in Ultimatum games. In the trust game analyzed here it is apparent that subjects with icons develop expectations and when they are not met, those subjects are willing to punish untrustworthy behavior.

Using the frequencies of trust, reciprocity, and punishment, we calculate the average payoff of subjects who choose a left move (trust) as an initial strategy, vs. those who choose a right move. In the sessions with facial icons, the average payoff to a trusting move was \$16.49; the payoff to moving right was \$14.97. Clearly, it paid to trust. In the sessions with nonfacial icons or blanks, the payoff to a left move was $13.76, and to a right move 15.45. In these sessions, trust was a riskier alternative.

Discussion.

These data are tantalizing, but the results are not completely clear. Facial icons appear to have a signaling effect -- at least with respect to reciprocated trust. Even in this highly controlled setting, however, the effect of smiles is not clear.

However, before drawing any conclusion from these data several caveats need to be registered. First there is an important concern about what is being signaled in this environment. In this setting subjects are randomly assigned an icon and subjects observe the icons of others. In other work (Eckel and Wilson, 1998; 1999) we find that subjects quickly "conform to type." That is they draw inferences about what their own icon means and then try to behave in line with expectations about their own icon. This compounds the empirical problem of determining what the counterpart's signal means to the subject. In the analysis presented here we have assumed away strategic signaling behavior on the part of the subject and have been concerned only with what signal might be inferred about the counterpart. Future research will disentangle these effects by limiting information about the player's own icon or that of the counterpart.

The second problem involves supergame play. This analysis assumes that subjects are treating each decision independently. However, in other work (Eckel and Wilson, 1999) we find clear effects of supergame play in which subjects spend a good deal of time using other games to signal a willingness to reciprocate or play cooperatively. Although these behaviors carry no specific signal for the individual, population characteristics do emerge. Given that there are only two types of icons in every population, this makes the inference problem a little easier for subjects as they play many periods. In order to deal with this problem, future research calls for one-shot games and more rigorous analysis of population dynamics within the repeated games.

Despite these concerns, we are intrigued by the fact that trust is reciprocated and that much of this reciprocated behavior is contingent on the second mover's icon holding facial characteristics. Although the numbers are small, it is encouraging to see that "punishment" for failing to reciprocate one's trust occurs more frequently for subjects with facial icon rather than those with no facial signals.

Conclusion

The ability to signal to and draw inferences about strangers is an important tool for facilitating trust and reciprocity in bargaining and exchange situations. In many bargaining settings there is a rich institutional structure that nurtures trusting behavior. Simple social mechanisms like gossip generate powerful tools for knowing whether a new partner can be trusted; at the same time such tools can pose powerful sanctioning mechanisms if one is deemed to be untrustworthy. Our work focuses on settings in which those mechanisms are absent. We are probing the foundations for drawing inferences about another, especially when there are gains from trust and trustworthy behavior. Our general sense is that people commonly resort to many devices to read something about the intentions of others.

There are many settings -- particularly in large urban areas -- in which the ability to judge the trustworthiness of strangers may be important. It seems clear that such inferences are drawn; however, the basis is unclear for the degree of trust and reciprocity that we can observe. Trusting behavior occurs with some frequency, often with strangers, and usually with that trust being reciprocated. This does not imply that all individuals are trusting to the same degree, or across all situations. Instead we think that trust is conditional and is rooted in individuals' ability to signal and read one another's intentions.

We have presented data from a series of surveys and experiments designed to provide preliminary evidence on the effect of facial expressions on trust and reciprocity. The results present some puzzles, but overall it seems clear that facial expressions are a potentially important influence on behavior of this type. In both sets of experiments presented here -- those using a human face and those using icons -- it is clear that individuals are choosing behavioral strategies at variance with the predictions of standard game theory. In addition, the use of facial icons seems a promising avenue for parsing the effect of facial expressions on behavior.

Subjects appear to be attuned to signals about their partners. A rich literature on facial expressions points to the importance of expressions for the communication of emotion. We think that it is easy to imagine that those expressions can serve to communicate intentions as well. The preliminary experiments reported here illustrate some of these tendencies. However, these results are not compelling, as confounding effects remain. In future research we will begin to systematically separate and examine these effects.

Bibliography

Adolphs, R., D. Tranel, H. Damasio and A. Damasio. 1994. "Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala." *Nature*. 372: 669-672.

Adolphs, R., Damasio, H., Tranel, D. and Damasio, A. 1996. “Cortical Systems for the Recognition of Emotion in Facial Expressions.” *Journal of Neuroscience.* 16 (23): 7678-7687.

Aggleton, John P. 1992. “The Functional Effects of Amygdala Lesions in Humans: A Comparison with Findings from Monkeys.” In John P. Aggleton (ed) The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction. New York: Wiley-Liss. (pp. 485-503)

Aggleton, J. P., and R. E. Passingham. 1981. "Syndrome produced by lesions of the amygdala in monkeys (*Macaca mulatta*)." *Journal of Comparative and Physiological Psychology* 95: 961-977.

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.

Ball, Sheryl and Catherine C. Eckel. 1998. "Stars Upon Thars: Status Differences in the Ultimatum Game". Working paper, Department of Economics, Virginia Polytechnic Institute and State University..

Ball, Sheryl, Catherine C. Eckel, Philip Grossman, and William Zame. 1998. "Status in Markets." Working paper, Department of Economics, Virginia Polytechnic Institute and State University.

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.

Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the Evolutionary Process.* Chicago: University of Chicago Press.

Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hyman, S. E., and Rosen, B.R. 1996. “Response and Habituation of the Human Amygdala During Visual Processing of Facial Expression.” *Neuron* 17 (5): 875-887.

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.

Carroll, James M. and James A. Russell. 1997. “Facial Expressions in Hollywood’s Portrayal of Emotion.” *Journal of Personality and Social Psychology.* 72 (1): 164-176

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 JohnTooby. 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.

Damasio, Antonio R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: Avon Books.

Darwin (1998) -- *The Expression of the Emotions in Man and Animals* (Series in Affective Science) by Charles Darwin. Paul Ekman (Editor) Oxford University Press.

Dawes, Robyn. 1980. "Social Dilemmas." *Annual Review of Psychology* 31: 169-193.

Deacon, Terrence W. 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York: W. W. Norton and Company.

Dimberg, U. and Ohman, A. 1996. "Behold the Wrath - Psychophysiological Responses to Facial Stimuli." *Motivation and Emotion* 20 (2): 149-182.

Dunbar, Robin. 1997. *Grooming, Gossip and the Evolution of Language.* Cambridge, MA: Harvard 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. 1998a. "Are Women Less Selfish Than Men?: Evidence from Dictator Games." *The Economic Journal*.

Eckel, Catherine C. and Philip Grossman. 1998b. "Chivalry and Solidarity: Gender Differences in Ultimatum Game Experiments." Working Paper, Department of Economics, Virginia Polytechnic Institute and State University.

Eckel, Catherine C. and Philip Grossman. In press. "Differences in the Economic Decisions of Men and Women: Experimental Evidence." *Handbook of Experimental Results*, edited by Charles Plott and Vernon Smith. New York: Elsevier.

Eckel, Catherine C. and Rick K. Wilson. 1998 "Social Signaling in Simple Bargaining Games." Paper presented at the Southern Economic Association Meetings, November 8-10, 1998, Baltimore, MD.

Eckel, Catherine C. and Rick K. Wilson. 1999. "Reciprocal Fairness and Social Signaling: Experiments with Limited Reputations." Paper Presented at the American Economic Association Meetings, January 3-6, 1999, New York, NY.

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 Unviersity 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, Paul, Richard J. Davidson, and Wallace V. Friesen. 1990. "The Duchenne Smile: Emotional Expression and Brain Physiology I." *Journal of Personality and Social Psychology*, 58 (2): 342-353.

Ekman, Paul and Wallace V. Friesen. 1978. *The Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, Paul and Wallace V. Friesen. 1982. "Felt, False, and Miserable Smiles." *Journal of Nonverbal Behavior*. 6: 238-252.

Ekman, Paul, Wallace. V. Friesen and Maureen O'Sullivan. 1988. "Smiles when lying." *Journal of Personality and Social Psychology*. 54: 414-420.

Ekman, Paul and Maureen O'Sullivan. 1991. "Who Can Catch a Liar?" *American Psychologist* 46(9): 913-920.

Elster, Jon. 1999. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge University Press (due out in March, 1999).

Fehr, Ernst, Georg Kircksteiger, and Arno Reidl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation. *Quarterly Journal of Economics* 108(2): 437-59.

Fernandez-Dols, J. M. and M. A. Ruiz-Belda. 1995. "Are smiles a sign of happiness -- Gold Medal winners at the Olympic games." *Journal of Personality and Social Psychology*. 69: 1113-1119.

Fernandez-Dols, J. M. and M. A. Ruiz-Belda. 1997. "Spontaneous Facial Behavior During Intense Emotional Episodes: Artistic Truth and Optical Truth." In James A. Russell and Jose Miguel Fernandez-Dols (eds.) The Psychology of Facial Expression. Cambridge: Cambridge University Press, pp. 255-274.

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.

Frank, Mark G. and Paul Ekman. 1997. "The Ability to Detect Deceit Generalizes Across Different Types of High-Stake Lies." *Journal of Personality and Social Psychology*. 72(6): 1429-1439.

Frank, Robert. 1988. *Passions Within Reason: The Strategic Role of the Emotions.* New York: W. W. Norton & Co.

Frank, Robert. 1997. "Nonverbal Communication and the Emergence of Moral Sentiments." In Ullica Segerstrale and Peter Molnar (eds.) *Nonverbal Communication: Where Nature Meets Culture*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, pp. 275-292.

Fridlund, Alan J., Ekman, Paul and Oster, Harriet. 1987. "Facial Expressions of Emotion: Review of Literature, 1970-1983." In Aron W. Siegman and Stanley Feldstein (eds.) *Nonverbal Behavior and Communication*, 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, pp. 143-224.

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.

Fudenberg, Drew and David K. Levine. 1992. "Maintaining a Reputation When Strategies Are Imperfectly Observed." Review of Economic Studies. 59 (3): 561-579.

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.

Hess, U., Banse, R., and Kappas, A. 1995. "The Intensity of Facial Expression Is Determined by Underlying Affective State and Social Situation." *Journal of Personality and Social Psychology* 69 (2): 280-288.

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*.

Izard, Carroll E. 1997. "Emotions and Facial Expressions: A Perspective from Differential Emotions Theory." In James A. Russell and Jose Miguel Fernandez-Dols (eds.) The Psychology of Facial Expression. Cambridge: Cambridge University Press, pp. 57-77.

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.

Kling, Arthur S. and Leslie A. Brothers. 1992. “The Amygdala and Social Behavior.” In John P. Aggleton (ed) The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction. New York: Wiley-Liss. (pp. 353-377).

Knutson, B. 1996. “Facial Expressions of Emotion Influence Interpersonal Trait Inferences.” *Journal of Nonverbal Behavior* 20 (3): 165-182.

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.

Ledyard, John O. 1995. “Public Goods: A Survey of Experimental Research.” In John Kagel and Alvin E. Roth (eds.), *Handbook of Experimental Economics*. Princeton University Press: Princeton, NJ.

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.

MacRae, C. Neil, Charles Stangor, and Miles Hewstone (Editors), *Stereotypes and Stereotyping*, Guilford Press, 1996

Morris, J. S., C. D. Frith, D. I. Perrett, D. Rowland, A. W. Young, A. J. Calder and R. J. Dolan. 1996. "A differntial neural response in the human amygdala to fearful and happy facial expressions." *Nature* (31 October 1996), V. 383: 812-815 **.**

O’Connell, Sanjida. 1998. *Mindreading: An Investigation into How We Learn to Love and Lie.* New York: Doubleday.

Ostrom, Elinor. 1998. “A Behavioral Approach to the Rational Choice Theory of Collective Action.” American Political Science Review. 92: 1-22.

Parkinson, B. 1996. “Emotions Are Social.” *British Journal of Psychology*. 87 (NOV): 663-683.

Pinker, Steven. 1994. *The Language Instinct.* New York: Harper-Collins.

Pinker, Steven. 1997. *How the Mind Works*. New York: W. W. Norton & Co.

Rapoport, Anatol. 1988. "Experiments with N-Person Social Traps I: Prisoner's Dilemma, Weak Prisoner's Dilemma, Volunteer's Dilemma, and Largest Number." *Journal of Conflict Resolution* 32(3): 457-72.

Rapoport, Anatol, and Albert M. Chammah. 1965. *Prisoner's dilemma; a study in conflict and cooperation.* Ann Arbor: University of Michigan Press.

Remnick, David. 1998. *King of the World : Muhammad Ali and the Rise of an American Hero.* New York: Random House.

Russell, J. A. 1993. "Forced-Choice Response Format in the Study of Facial Expression." *Motivation and Emotion*. 17: 41-51.

Russell, J. A. 1994. "Is there universal recognition of emotion from facial expression?" *Psychological Bulletin*. 115: 102-141.

Sally, David. 1995. "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992." *Rationality and Society* 7: 58-92.

Smith, Vernon L. 1998. "The Two Faces of Adam Smith." *Southern Economic Journal* 65(1): 1-19.

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. 33-47 in W. G. Austin and S. Worchel (eds.) *The Social Psychology of Intergroup Relations*. Monterey, 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 the Social Psychology of Intergroup Relations*. London: Academic Press Inc. (1978)

Weiskrantz, L. 1956. "Behavioral changes associated with ablations of the amygloid complex in monkeys." *Journal of Comparative and Physiological Psychology* 49: 381-391.

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.

Young, A.W., Hellawell, D. J., Vandewal, C., and Johnson, M. 1996. "Facial Expression Processing After Amygdalotomy." *Neuropsychologia* 34 (1): 31-39.

Zebrowitz, Leslie A. 1997. *Reading Faces: Window to the Soul?* Boulder, CO.: Westview Press.

Table 1
Percentage Choosing an Item
(Frequencies in Parentheses)

Item:	***Treatment Stimulus***				
	Blank	*Male Smile*	*Male Neutral*	*Female Smile*	*Female Neutral*
Left Move (Trust)	39.5 *(30)*	45.7 *(42)*	28.4 *(31)*	38.7 *(24)*	53.6 *(37)*
Image is Friendly	59.2 *(45)*	51.1 *(47)*	10.1 *(11)*	69.4 *(43)*	7.5 *(5)*
Image is Cooperative	63.2 *(48)*	52.2 *(48)*	15.6 *(17)*	53.2 *(33)*	21.7 *(15)*
Total n	*76*	*92*	*109*	*62*	*69*

Table 2

Probit Estimates of Choosing an Initial Trust Move

Dependent Variable = 1 if a Trust move (left) is taken

(Standard Errors in Parentheses)

Variable	***Model 1***	***Model 2***	***Model 3***
Intercept	-.227* (.*146*)	-.267* (.*146*)	-.267* (.*146*)
Face	.023 (.*161*)	-.140 (.*184*)	-.303 (.*194*)
Female	--	.263* (.*142*)	.661*** (.*199*)
Smile	--	.120 (.*139*)	.461** (.*183*)
*Female*Smile*	--	--	-.838*** (.*287*)
Log Likelihood *Number of observations*	-274.90 n=408	-272.80 n=408	-268.51 n=408

*p<.10, **p<.05, ***p<.01

Table 3
Effect of Facial Characteristics on Perceptions of Icons
Standard Errors in Parentheses

	Model 1		**Model 2**	
	Behavior	**Affect**	**Behavior**	**Affect**
Intercept	4.11*** (.10)	4.10*** (.10)	4.15*** (.14)	4.05*** (.14)
 **SMILE**	-.09 (.10)	-.78*** (.10)	-.34* (.19)	-.96*** (.18)
 **UPBROW**	-.80*** (.11)	-.57*** (.11)	-.67*** (.18)	-.42** (.18)
 **FROWN**	.13 (.10)	.70*** (.10)	.27 (.19)	1.04*** (.18)
 **DOWNBROW**	1.02*** (.10)	1.14*** (.10)	.87*** (.16)	1.17*** (.16)
SEX (1=Female)	.05 (.09)	.19** (.08)	.04 (.08)	.18** (.08)
 **Interaction**			-.21 (.26)	-.09 (.26)
 **Interaction**			-.33 (.24)	-.69*** (.24)
 **Interaction**			.88*** (.24)	.62*** (.24)
 **Interaction**			-.17 (.26)	-.35 (.26)
r2	.40	.51	.45	.54
	*p<.10	**p<.05	***p < .01	

Table 4
Percentage of First-Movers Choosing an Initial Trusting Move (Left)
(frequencies in italics)

	Devious	Angry	Happy	Sad	Rect.	Oval	Blank	Total
Devious	80.0 *16/20*	--	55.6 *10/18*	70.0 *7/10*	--	--	--	68.8 *33/48*
Angry	--	75.0 *3/4*	66.7 *6/9*	--	--	--	--	69.2 *9/13*
Happy	80.0 *12/15*	0.0 *0/1*	66.7 *12/18*	--	--	--	--	70.6 *24/34*
Sad	50.0 *2/4*	--	--	100 *4/4*	--	--	--	75.0 *6/8*
Rect.	--	--	--	--	60.0 *6/10*	66.7 *10/15*	--	64.0 *16/25*
Oval	--	--	--	--	44.4 *4/9*	57.1 *4/7*	--	50.0 *8/16*
Blank	--	--	--	--	--	--	56.25 *18/32*	56.25 *18/32*
Total	76.9 *30/39*	60.0 *3/5*	62.2 *28/45*		52.6 *10/19*	63.6 *14/22*	56.25 *18/32*	64.8 *114/176*

Table 5

Probit Estimates for First Move in Game 2

(Probability of Moving Left -- Standard Errors in Parentheses)

Variable	***Model 1***	***Model 2***	***Model 3***
Intercept	.423* (.252)	.411 (.303)	.435 (.254)
Blank	-.069 (.299)	-.064 (.342)	-.069 (.299)
Face	.274 (.239)	.147 (.311)	-.102 (.610)
Period	-.014 (.010)	-.013 (.010)	-.015 (.010)
Paired (No Expression)	--	.011 (.403)	--
Paired (Expression)	--	.313 (.266)	--
Smile	--	--	.596 (.616)
Eyebrows	--	--	.630 (.683)
*Smile*Eyebrows*	--	--	-1.055 (.743)
Log Likelihood *Number of subjects*	-111.85 n = 176	-111.15 n = 176	-110.33 n = 176

*p<.10, **p<.05, ***p<.01

Table 6

Probit Estimates for Reciprocal Move in Game 2

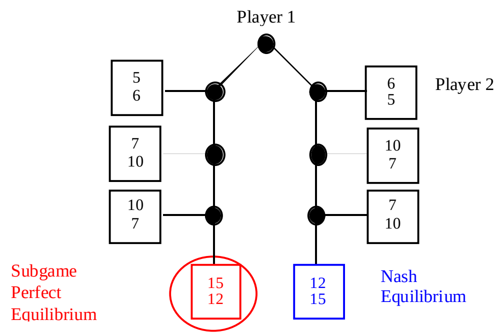
(Probability of Reciprocating a Trusting Move -- Standard Errors in Parentheses)

Variable	***Model 1***	***Model 2***	***Model 3***
Intercept	.044 (.312)	-.209 (.385)	.036 (.313)
Blank	.551 (.400)	.825* (.464)	.550 (.400)
Face	.911*** (.307)	1.269*** (.422)	1.018* (.502)
Period	-.012 (.014)	-.013 (.014)	-.011 (.014)
Paired (No Expression)	--	.636 (.531)	--
Paired (Expression)	--	-.173 (.337)	--
Smile (of first-mover)	--	--	-0.356 (.464)
Eyebrows (of first-mover)	--	--	.509 (.362)
Log Likelihood *Number of subjects*	-64.50 n = 114	-63.64 n = 114	-62.63 n = 114

*p<.10, **p<.05, ***p<.01

Figure 1
Sample Sequential Games

Game A



Game B

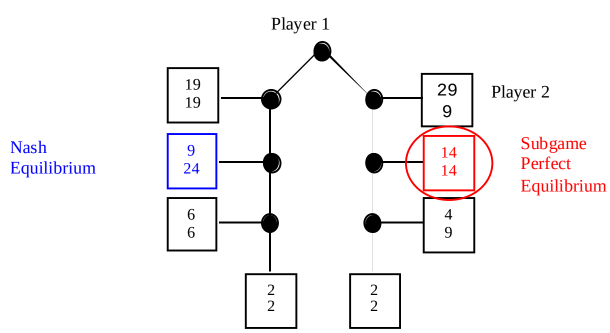


Figure 2
Mean Ratings of Faces (Standard Deviations in Parentheses)

Male Smile

Happy	1.62 (.64)
Sad	4.29 (.66)
Angry	4.27 (.68)

Male Neutral

Happy	3.61 (.73)
Sad	2.75 (.83)
Angry	3.00 (.75)

Female Smile

Happy	1.53 (.60)
Sad	4.37 (.63)
Angry	4.31 (.70)

Female Neutral

Happy	3.89 (.69)
Sad	2.73 (.78)
Angry	2.67 (.92)

Figure 3
Game for Experiment 1

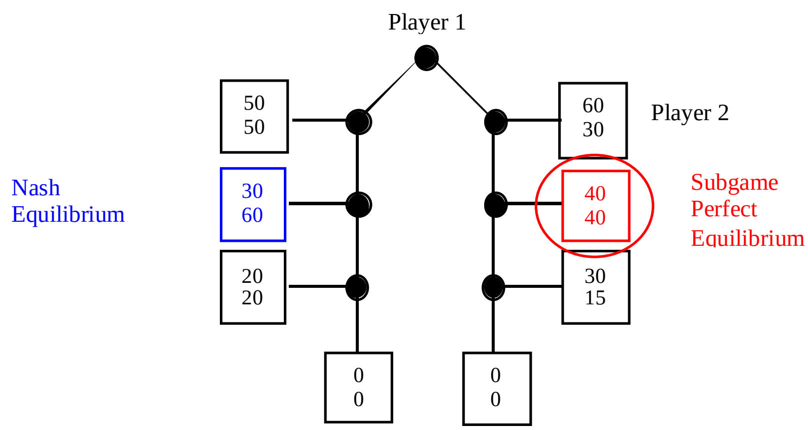


Figure 4
Icons Used in Survey

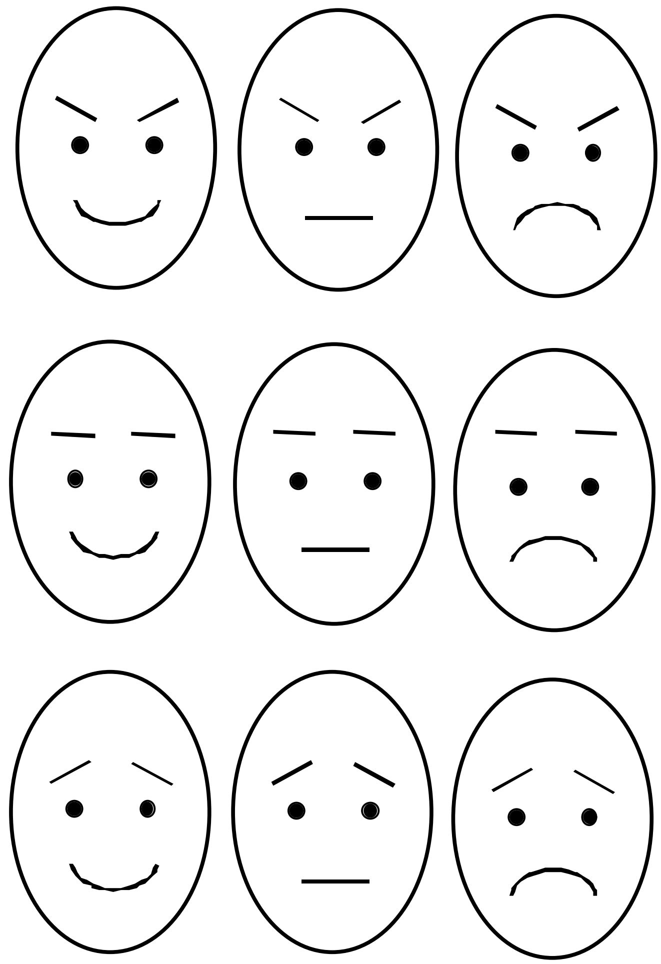


Figure 5
Mean Ratings of Icons on Semantic Differential Items

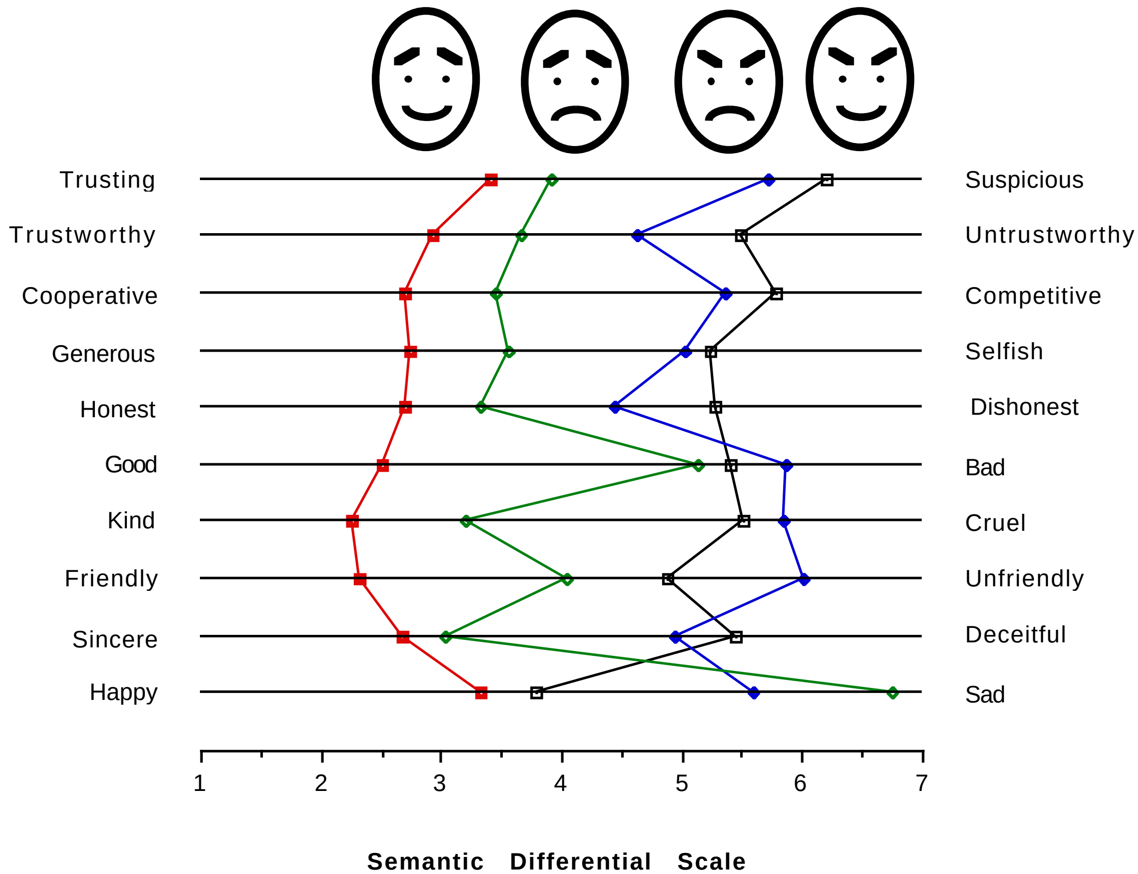


Figure 6
Icon Pairs Used in Experiment

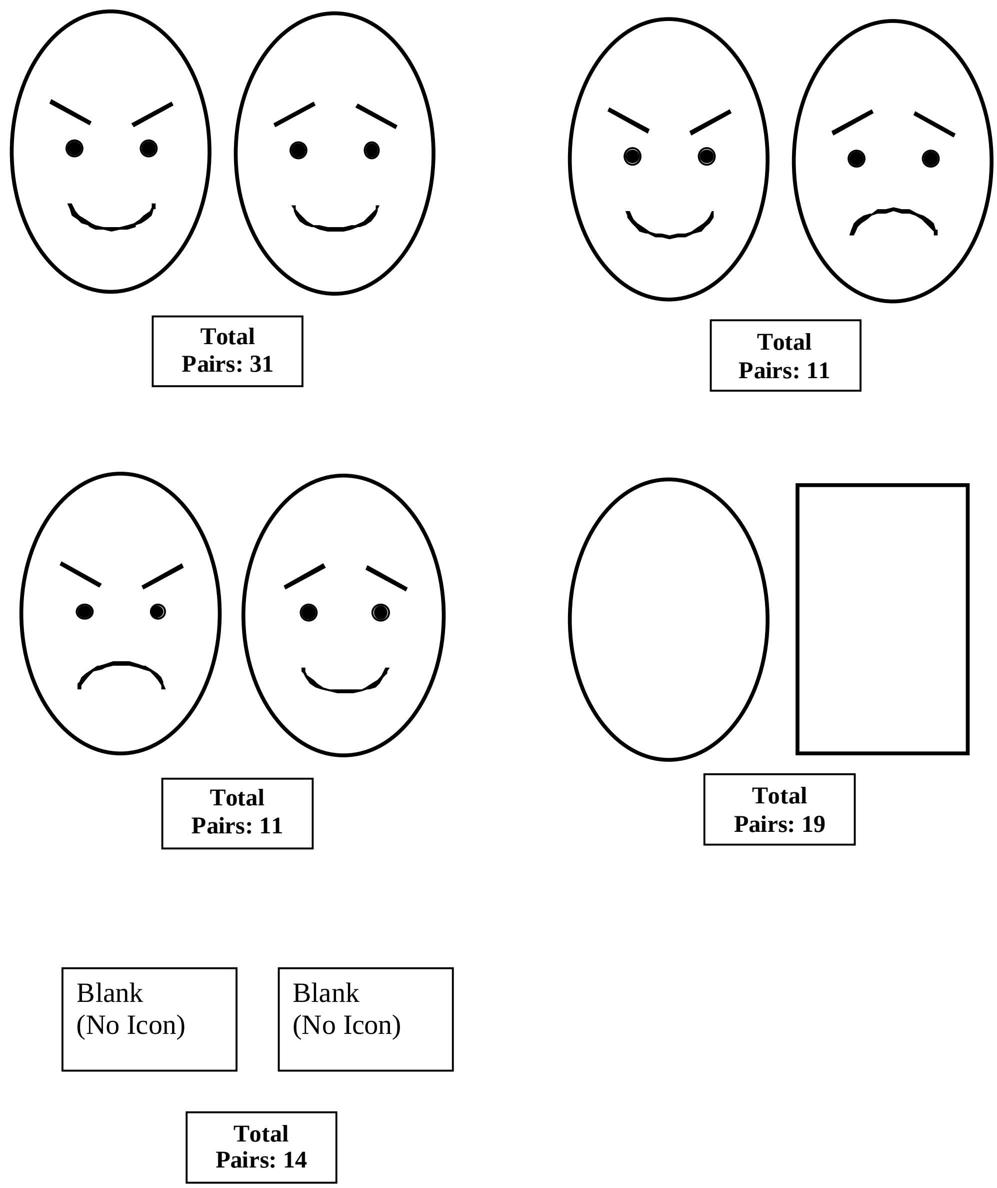
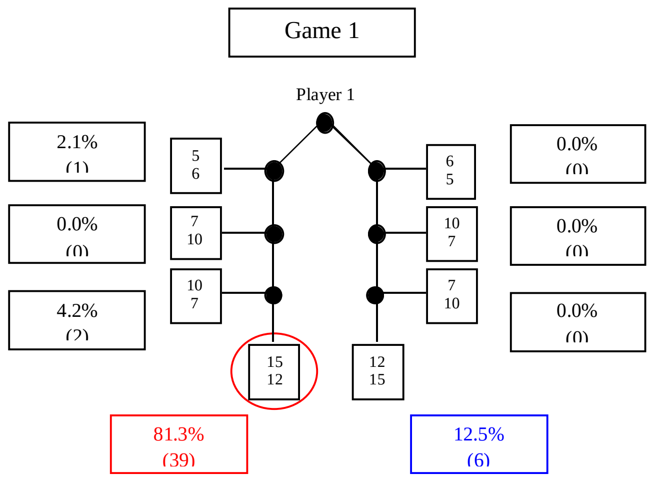
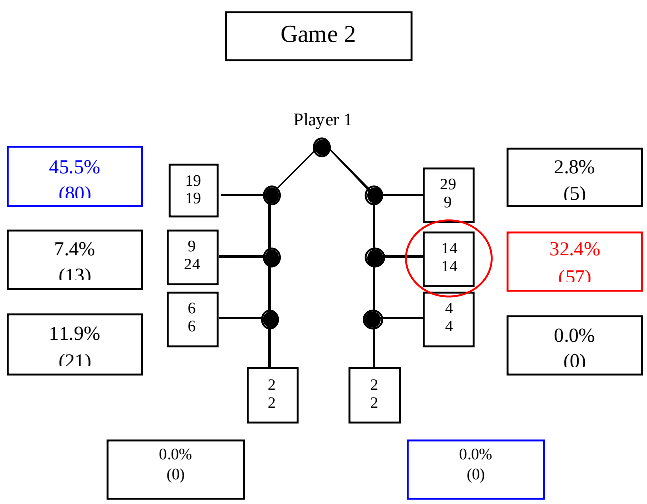


Figure 7
Distribution of Outcomes in the Experiment





Appendix 1
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.

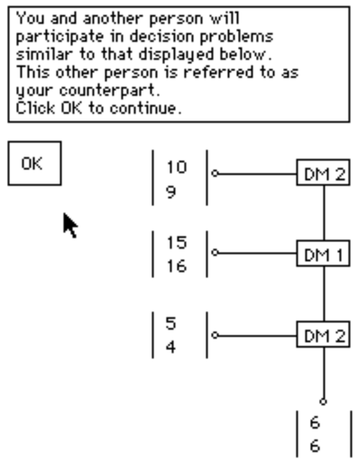
On the sheet of paper I have provided, please record your potential earnings for each decision. This will help you keep track of what you earn at the end.

Click OK when you are ready to continue.

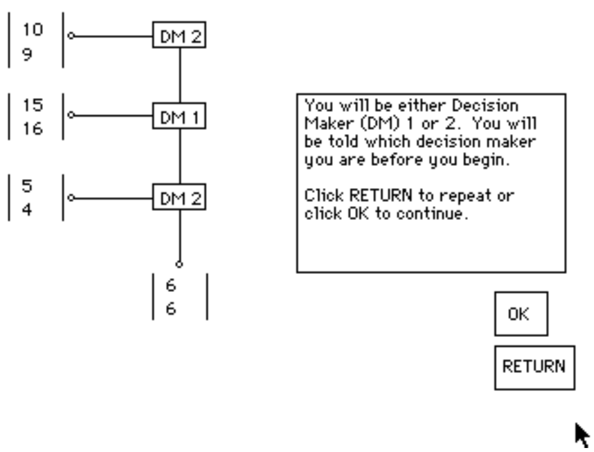
OK

-

Screen 3

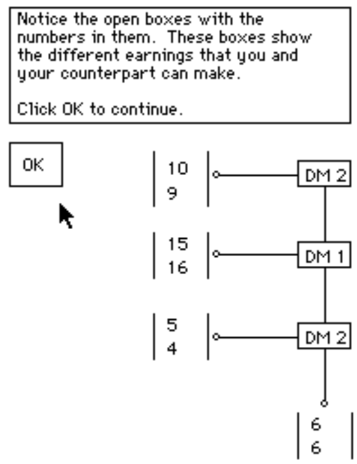


Screen 4

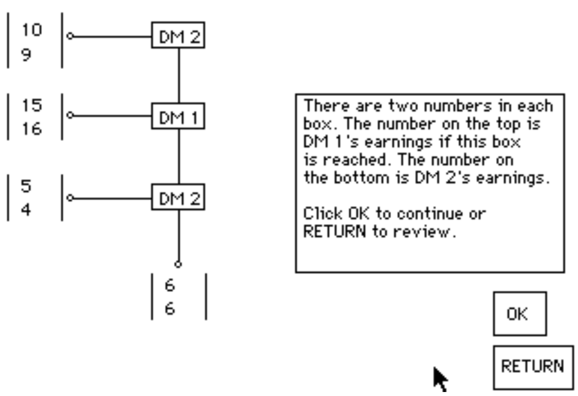


-

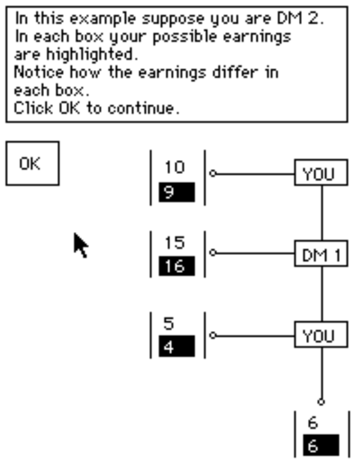
Screen 5



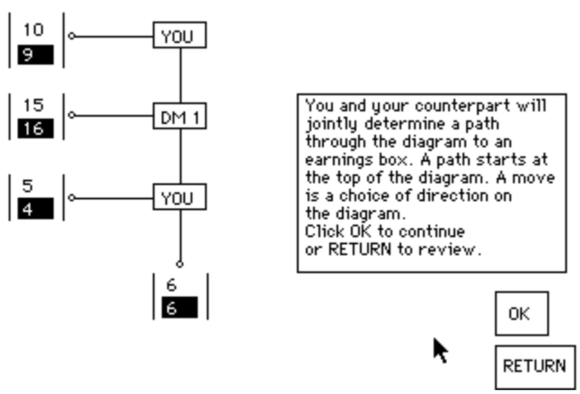
Screen 6



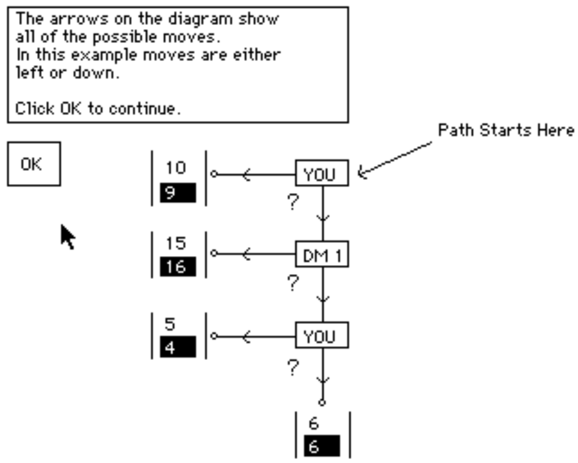
Screen 7



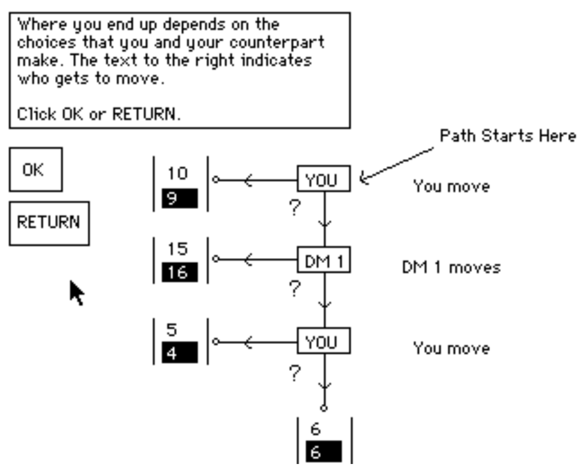
Screen 8



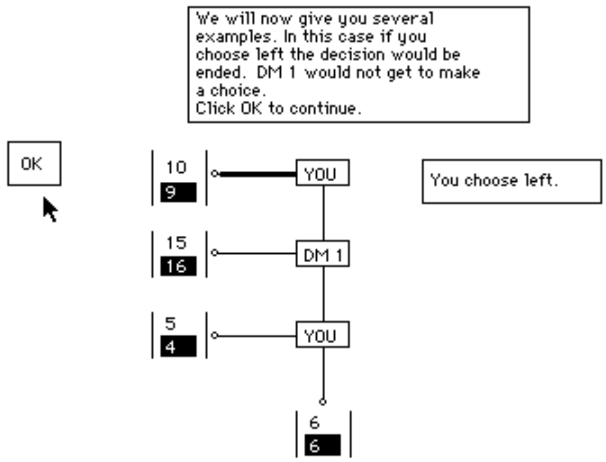
Screen 9



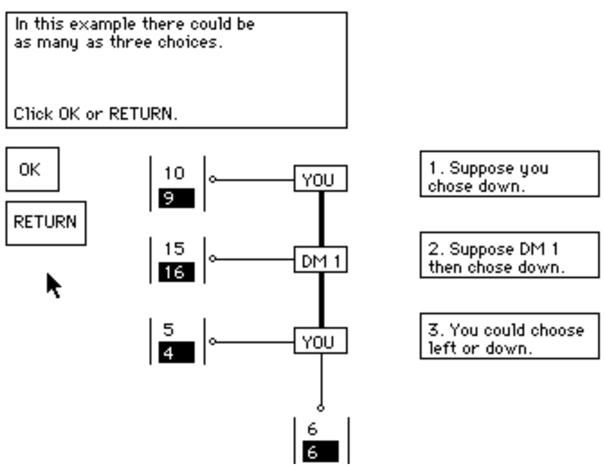
Screen 10



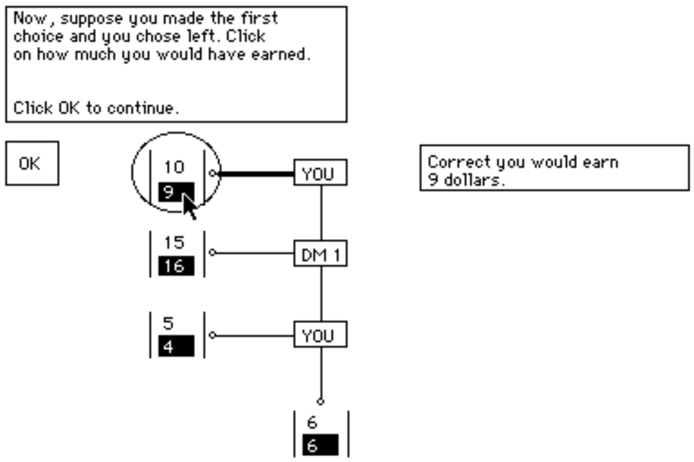
Screen 11



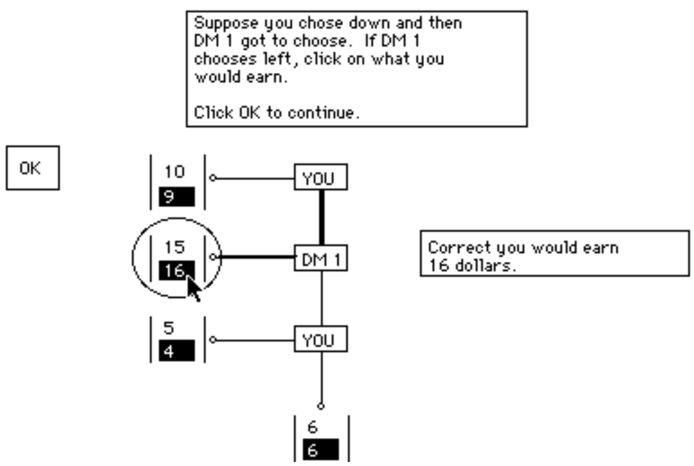
Screen 12



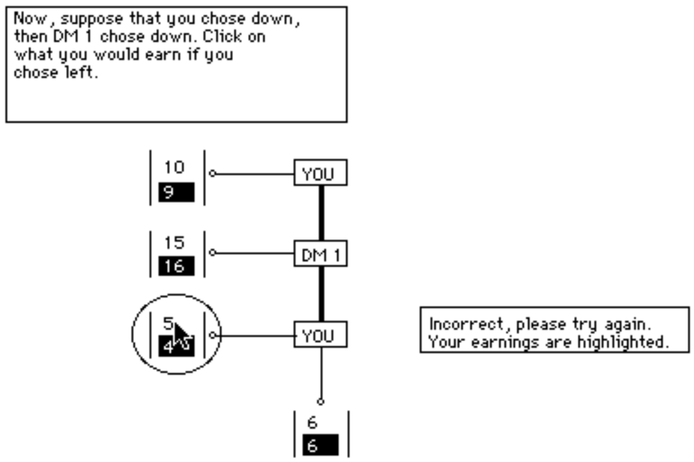
Screen 13



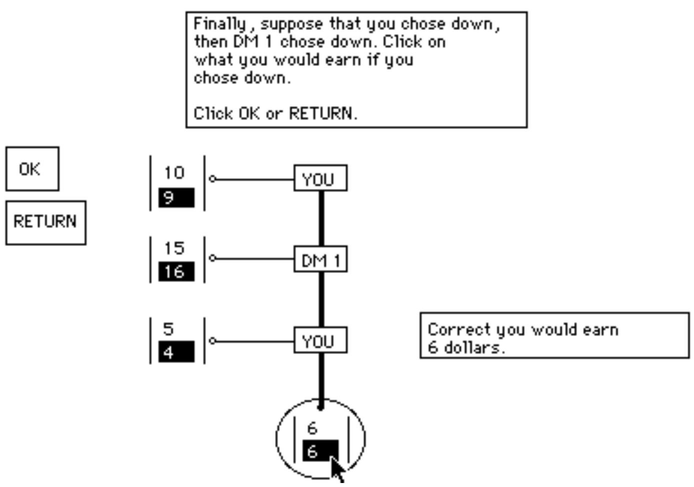
Screen 14



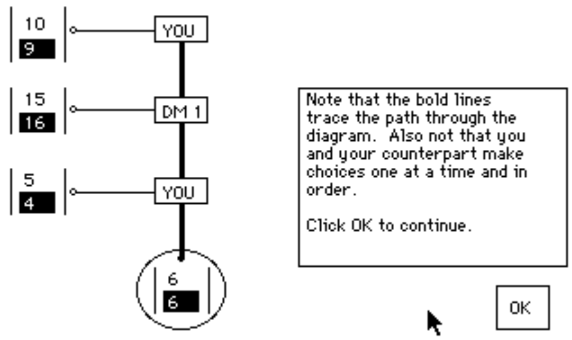
Screen 15



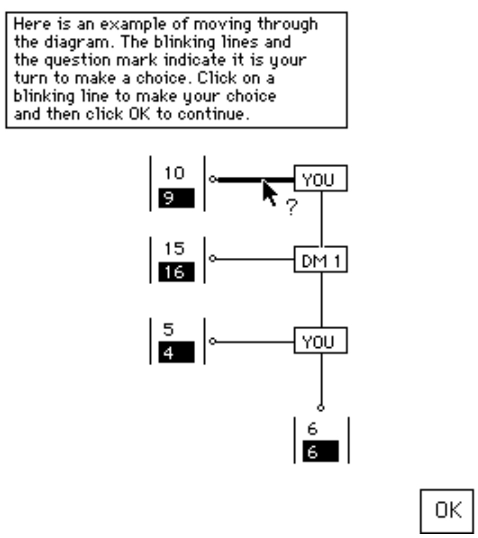
Screen 16



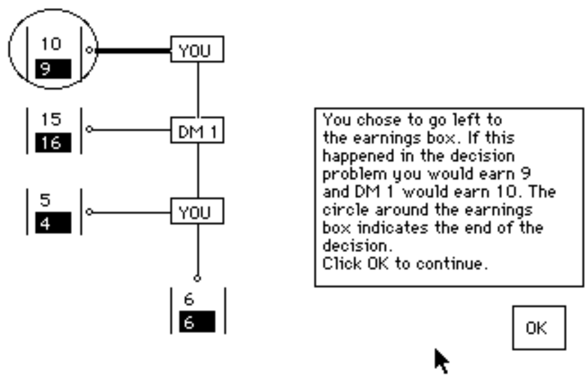
Screen 17



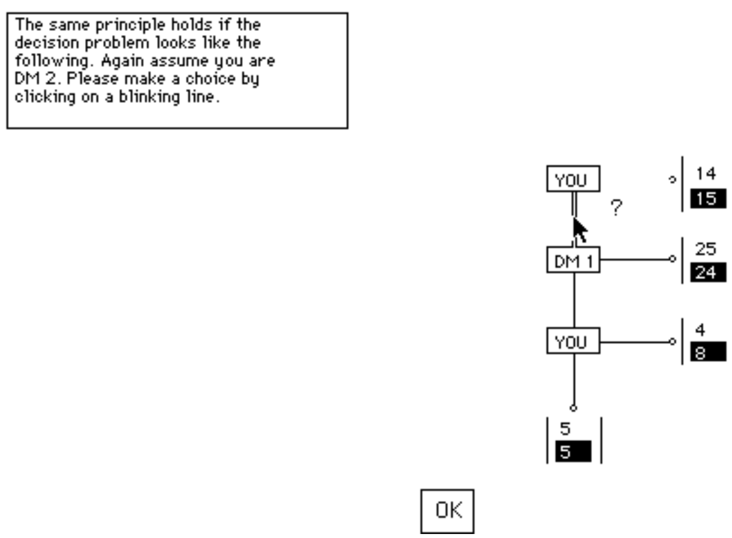
Screen 18



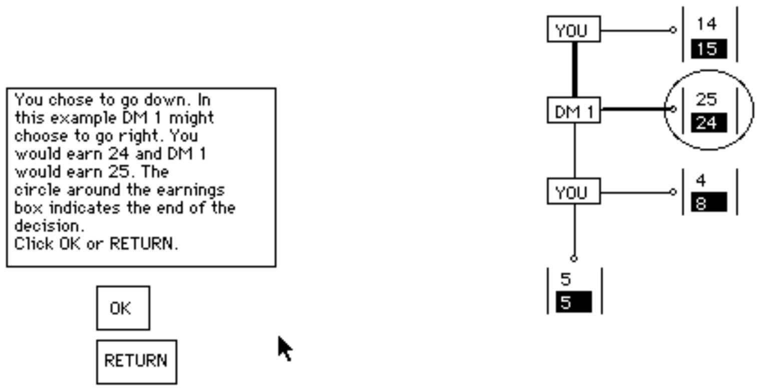
Screen 19



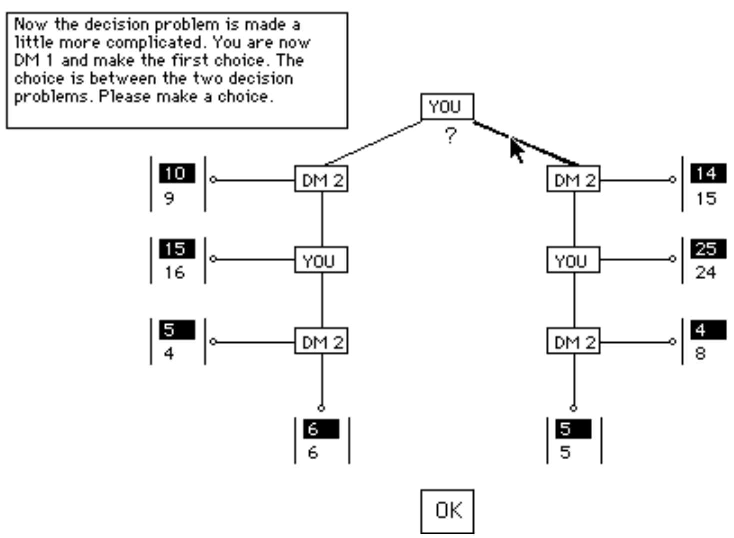
Screen 20



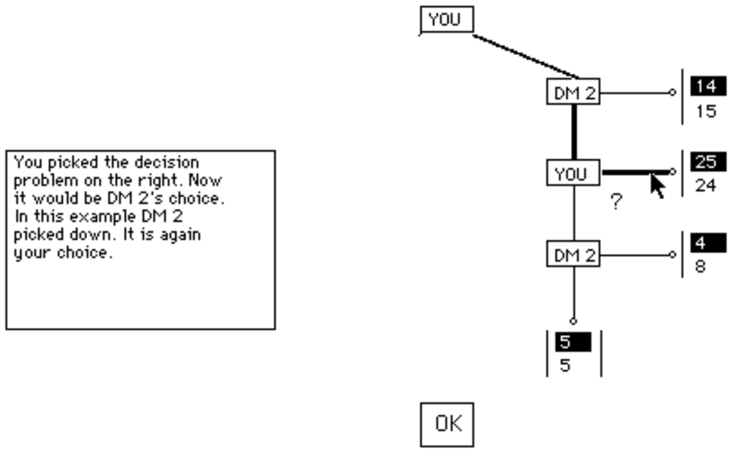
Screen 21



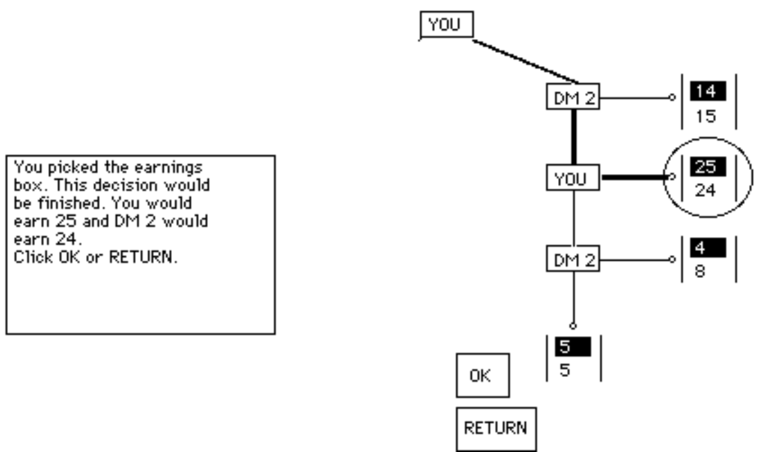
Screen 22



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