The Human Face of Game Theory

by

Catherine Eckel
Virginia Polytechnic Institute & State University
National Science Foundation

and

Rick K. Wilson
Rice University
National Science Foundation

The authors gratefully acknowledge support by the National Science Foundation. The agency bears no responsibility for any conclusions reached in this research. Valuable assistance for the pilot 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. Finally, we spent some time bending the ears of Betsy Hoffman, Kevin McCabe and Vernon Smith. No doubt we’ll be doing it again. Paper prepared for presentation at the annual meeting of the Economic Science Association, Tucson, AZ, September 19-20, 1997.
The Human Face of Game Theory

Introduction

Much of game theory rests on the foundation of common knowledge: actors sharing common beliefs about one another and about the nature of the game being played. While common knowledge is reflexively invoked, it is rarely realized. Whether actors examine a game, examine one another, and then arrive at a similar set of beliefs remains an open question.

Drawing on concepts from evolutionary psychology, this paper focuses on a social cue that can affect the beliefs held by actors playing a simple game. Our contention is that the image of a player’s counterpart contains information that is used by the player in formulating beliefs and subsequent actions. The population of possible players encompasses many different types, whose personal characteristics may be correlated with play and whose image may signal a type.

Players choose strategies contingent on expectations about the behavior of their counterpart. Part of those expectations are based on past experiences and part are based on inferences about potential patterns of play. With respect to the latter, it is clear that both players embody characteristics (such as gender, age, socio-economic status) that signal their types. Moreover, each can choose to display additional social signals in the form of facial expression, attire, or demeanor. These inherent and intentional signals provide information to the players about their counterparts and the initial strategies they are likely to choose. As play is repeated, observed strategy choices allow players to update their beliefs, and more accurately judge a player’s type. Initial play, however, is important in determining the path of play and, particularly in a multi-equilibrium game, have a substantial impact on the equilibrium that is selected.

Imagine two individuals are facing one another for the first time in some form of social exchange. They know nothing about one another, but each has to make a decision and those decisions will jointly affect their payoffs. How does each anticipate the other’s actions? How does an agent judge whether a partner is trustworthy, or if reciprocity can be expected? At a very basic level, should they begin by making a cooperative or competitive move? What each subject does at the outset may markedly affect subsequent interactions, and as such, first moves are always important.

In this paper we present preliminary findings from experiments designed to test the effect of social signals on initial play. We draw heavily on a design previously used by Hoffman, McCabe and Smith (1996) and McCabe, Rassenti and Smith (1997).
Prior Research

We are interested in the ways in which evolutionary psychology informs a set of key issues in game theory. In particular, how do actors draw inferences about their counterparts in social, economic and political interactions -- especially at the outset? In modeling repeated interaction, game theory has focused on reputation and reputation-formation, but has had a difficult time characterizing initial interactions. We examine initial play, which is based on actors’ homegrown priors. We argue that these arise in turn from a kind of population reputation inferred from the characteristics of a player. Important questions remain concerning the origin and accuracy of those reputational priors. Evolutionary psychology can be used to develop hypotheses about the ways in which humans draw inferences about others.

We start from the idea that there are evolutionarily-derived mental modules used by humans to make decisions. That is, instead of human beings having a single, general purpose computational mechanism (a mind) that uses general-purpose rules to solve problems, the brain is a disaggregated information processing mechanism that holds many different specialized processing capabilities. Several million years of evolutionary adaptation to solve the social-exchange problems involved in hunting and gathering have led to the development of mechanisms of social cognition that are useful in dealing with social exchange problems. Evolutionary psychologists argue that the collective benefits of cooperation have led to the selection of mechanisms that promote cooperative behavior, particularly in the form of reciprocal altruism. De Waal (1996) describes such mechanisms for reciprocity even in nonhuman primates.

Imagine a human being with at least two mental modules for determining interactions with others. Hoffman, McCabe, and Smith (1996) (HMS) and McCabe, Rassenti and Smith (1997) (MRS), taking a page from Cosmides and Tooby (1992), have characterized such a mechanism as a “friend or foe” detector. That is, when confronting another, an actor slips into either a cooperative or a competitive mental module. Such an approach has the advantage of helping to explain the higher-than-expected, but varied, amount of reciprocal behavior observed in controlled laboratory experiments. A key piece of the puzzle, however, is an understanding of what triggers a particular distinct mental module.

There is a substantial literature noting that humans, primates, and many other mammals use facial expressions (as well as other posturing) to signal particular states of mind. Indeed, Charles Darwin early on published a treatise on the subject entitled *Expression of the Emotions in Man and Animals* (1872). Much of the current literature,
especially for humans, focuses on the emotional content of facial expressions. The central finding is that, cross culturally, facial expressions carry the same message. A smile appears to be a universal signal triggering a pleasant emotional response. Facial expressions, then, activate emotional affect in the target. Fridlund (1994) argues that facial expressions are always social signals, designed (by evolution) to elicit a particular response from an observer. The behavioral ramifications of emotional affect are largely unexplored, however.

Evolutionary psychologists have proposed 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.

Such a fanciful reconstruction is not without empirical support. 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 interacted 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. 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 addition there is a good deal of suggestive evidence that facial cues are fundamental for cementing social relations. In an imaginative study Johnson et al. (1991) traced the reaction of new born infants to a paper stimulus about the size and shape of a human head. A variety of stimuli were 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 found that newborns paid 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. In follow-up experiments with the same children, the researchers discovered that by the age of five months infants tracked all of the images at roughly the same rates. These researchers conclude that children arrive at birth with a system that first orients them toward face-like patterns and that, second, they develop a more mature 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 circuity that will subsequently underlie face-processing in the adult." (p. 18).

Finally, Antonio Damasio (1994) reviews a substantial body of neuro-physiological literature pointing to the importance of emotional response for rational behavior. While Damasio is very concerned with individual decision making, he spends a great deal of time worrying about how social signals are processed by individuals with brain damage. His touchstone is Phineas Gage, a construction foreman who, in 1848, suffered an accident that severely damaged his prefrontal lobe. Despite suffering tremendous damage to his brain, Gage was able to function as an ordinary being. He was able to walk, talk and do all of the things that ordinary human beings are able to do. Indeed, following his accident he lived for another 13 years. However, Gage was unable to read the emotional content of others and was often unable to make decisions. This often led him to exhibit inappropriate behavior and to fail at most basic social activities. Damasio uses Gage as an exemplar to illustrate the importance of particular regions of brain for processing social cues.

Economists typically model exchange as anonymous and the outcome of the exchange is assumed to be independent of the identities of the actors. But much economic interaction occurs between individuals who know, or at least observe, each other. The question naturally arises then, if exchange is social, how is the exchange affected if a particular mental module is activated by the characteristics of the participants?

If facial expressions activate different modules, then this may give great insight into one psychological process that leads to spontaneous cooperative (reciprocal) behavior in humans. Our approach attempts to manipulate the reputational prior inferred by experimental subjects. We examine the effect of the signals embodied in facial expressions on the behavior of agents in games with financial stakes.
Design

We report two experiments that share the same design. The first is a pilot experiment conducted with large undergraduate classes. To motivate subjects, a small number were randomly selected from the class and paid for their decisions. The second experiment was conducted in a computer controlled setting, and had a smaller number of subjects, all of whom were paid for their participation.

The design partially replicates that found in McCabe, Rassenti and Smith (1997 -- hereafter MRS). In their experiment a group of subjects plays the extensive form game found 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.

As MRS argue, this game has special features that permit inferences about which of two mental modules is activated for Player 1. 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.

In the subgame represented by the right branch, there is a Nash equilibrium (also the Subgame Perfect Nash equilibrium for the full game) at (40,40), where Player 2 chooses Down and Player 1 chooses Right. The left-branch subgame contains the symmetric joint maximum (50,50). But if both players are maximizing their own payoffs, Player 2 will choose Down, then Player 1 will choose left, reaching the Nash equilibrium of the left-branch subgame at (30,60). This subgame also incorporates an opportunity for Player 1 to punish Player 2, at some cost, by choosing Down, though this strategy is payoff-dominated.

If Player 1 expects payoff-maximizing play from Player 2, he will choose Right at the first node of the game. A decision to play Left indicates that Player 1 expects Player 2 to reciprocate by choosing Left, reaching (50,50). If Player 2 fails to comply, Player 1 can then punish her by choosing Down.

This game has the virtue of containing an initial move in which one subject can enter either a competitive game (the right branch) or a reciprocal game with a cooperative payoff (the left branch). Our conjecture (and that by HMS and MRS) is that the full play of each branch requires that subjects activate a mental module for either reciprocal or competitive calculation.
Our experiments replicate the instructions and general layout of the MRS experiment, but there are several differences. In the pilot experiments an entire classroom of subjects was given the instructions by the same experimenter (see the Appendix), 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 were programmed into the computer. (Subjects were deceived in this respect.)

At the conclusion of the instructions, depending on the treatment variable, subjects were shown an image of their counterpart and then a sheet was handed out with the decision problem. That sheet resembled Figure 1 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 approximately five percent 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 experiments.

In the second experiment subjects who had just completed a public goods game were asked if any would like to volunteer for a different experiment that would take place immediately.¹ In the public goods game, the subjects participated in the same group; now they were told they would participate with counterparts at a remote site. Once again, deception was used: subjects in fact played against the computer. The computer was programmed to replicate strategies resembling the distribution of choices reported in the MRS experiment. Subjects were informed that they were participating in a pilot experiment and that their counterpart was in Arizona. To heighten the sense that subjects were playing against an opponent at a remote site, once a choice was made, subjects were told to wait for the other participant to make a decision. That wait was randomly set between 30 and 130 seconds. When their decisions were concluded, subjects completed a brief questionnaire and manipulation check.

¹ These experiments were conducted at two separate locations. The pilot experiments in the classrooms were all conducted in large introductory economics classes at Virginia Polytechnic Institute and State University. The computer controlled experiments were run at Texas A&M.
Deception was critical to the experiment. The primary manipulation involved providing a fixed cue to each subject in the form of a facial image. Subjects were shown one of three cues shown in Figure 2. Three posed images were selected that were thought to capture three distinct emotional states. All three images depicted the same male undergraduate. The images were digitally recorded and then cropped to be the same size, and reflected happy, neutral and angry expressions. A fourth group was also run in which no image was shown to subjects. This control group is directly comparable to the MRS one-shot experiments. In the pilot experiment the image was displayed on a projector (either from a computer screen or an overhead slide), while in the computer-controlled experiment the image was displayed on the computer screen.

A second difference is that our game does not contain the "exit option" which appears at the very top of the MRS decision tree. MRS informed us that subjects never chose this option, and we removed it to simplify the structure of the game.

A third difference, which applies only to the computer controlled design, is that the left and right branches of the game shown in Figure 1 were randomized. Under the MRS design a left move and a right move always correspond to the branches shown in Figure 1. In our design, to control for a possible left/right bias among subjects (or the population), the branches were randomly assigned to either left or right. To minimize confusion in the discussion, we always report choices as they conform to figure 1.

Subjects only made a single choice in this experiment. They were paid in francs at a conversion rate of $.10 per franc. It took a subject approximately 15 minutes to complete the task and they made, on average, $4.05 for this experiment. Following the conclusion of this bargaining game subjects were debriefed. The subjects in the pilot experiment were given additional information about the structure of the game and told that the aim of the research was to understand social signaling. Those subjects in the computer controlled setting were told they were not actually playing against another subject at a computer laboratory at the University of Arizona. Instead they were thanked for their participation and told they were helping the experimenter design a new experiment. While they had been playing against a computer, that computer was programmed to reflect the decisions made by other people in this same decision setting.

**Analysis**

Our principal conjecture is that "friendly" social cues are critically important for inducing a cooperative move. MRS recognize this point in their experiment using the
same game. They argue at length that subjects opt for a cooperative move because they consider their own beliefs about the distribution of reciprocal tendencies among the population and respond accordingly. Rather than always choosing the "right" branch in the game given in Figure 1 (which is the non-cooperative game-theoretic prediction), their subjects choose "left" 50 percent of the time in the single-shot game. Using their results as a baseline, we first replicate their findings and then compare what happens when different types of social signals are introduced. Instead of relying on conjectures about the population, subjects are given the image of a specific individual to make inferences about.

Experiment 1 -- Pilot Test.

Unlike the MRS findings, our pilot experiment shows that subjects are far less likely to choose left. The first row for Table 1 reports the percentage of subjects choosing left at the first move. The findings are quite striking. Where subjects were given no image they selected the left branch 39.5 percent of the time, significantly less than what MRS report.

By comparison, when subjects were given a visual cue of their counterpart, their behavior shifts markedly. The percentage choosing left climbs to 44.7 percent when subjects were presented with an image with a smile. For both neutral and angry images the percent of subjects choosing a cooperative moves declines precipitously to roughly 28 percent of the total. The differences between the smile and neutral or angry images are significant under a chi-square test \( \chi^2 = 6.936, p = .008 \) smile/neutral and \( \chi^2 = 5.124, p = .024 \) smile/angry). The smile image results in far more cooperative behavior than under the neutral or angry image manipulations. These differences are marginally significant from the no-image, control condition under simple chi-square tests \( \chi^2 = 2.968, p = .085 \) for the control/neutral pairing and \( \chi^2 = 2.186, p = .139 \) for the control/angry pairing. Two things are apparent from these pilot test data. First, the degree to which subjects choose a cooperative strategy (a left branch first move) is substantially lower than that found in MRS. Second, when presented with images of their counterpart subjects alter their behavior.

Following their initial decision -- whether to choose left or right -- subjects were asked a series of questions about their counterpart in the experiment. Because this is a between-group design, subjects in each group only rated one image. Nonetheless, these data enable us to determine whether subjects perceived differences between the image stimuli. Two questions are of special concern: whether subjects felt their counterpart
was happy or sad, and whether subjects felt their counterpart was cooperative or uncooperative. Both questions were asked on a three point scale, with a value of 1 tied to a happy/cooperative rating, 2 as a neutral rating and 3 as a sad/uncooperative rating. Table 1 reports the means for each of the groups. While these ratings are after-the-fact, nonetheless they show a clear ordering. Subjects viewing an image with a smile provide a more positive rating than subjects viewing a neutral image. In turn a neutral image is more positively rated than an angry image. Except for one, all of the ordered pairs are statistically significantly different at the .05 level using a t-test for both items. The paired comparison on the cooperative/uncooperative item is significantly different at p<.07. These orderings are consistent with the aggregate patterns of choosing the cooperative, left, branch. Subjects who observed an image with a smile rated the image more highly and were more likely to choose left than when presented images with a neutral or angry expression. While subjects in the angry image group rated their image less positively than those in the neutral image group, they do not differ in the rates in which they choose the cooperative (left) branch of the game.

We also asked subjects who had observed no image to rate their counterpart. These ratings are given in column one of Table 1. Subjects were given no information about their counterpart except that it was an individual who had previously participated in this experiment. These findings appear to be consistent with the MRS conjecture that subjects will draw cooperative, friendly inferences about their counterpart, even though they have no cues about that counterpart. In essence they may think their counterpart is "just like them."

Finally, we asked subjects whether they thought the instructions used in the experiment were clear. Only 64.3 percent thought the instructions were clear. This leads us to wonder whether subjects understood the nature of the game. This poses a potential source of concern for this experimental design. Interestingly, however, when controlling for a subject's self-reported comprehension, those reporting the instructions as clear chose the left branch at rates similar to those who reported the instructions were unclear.

**Experiment 2 -- Computer Controlled.**

It may be that in the pilot experiment subjects doubted whether they were paired with someone who had previously made a choice. Although this point was emphasized in the instructions, subjects may have been skeptical. In the second experiment subjects were led to believe that they were participating with another subject at a remote site in real-time. They had just completed an experiment in which they knew they were participating with others in their group. They had little reason to believe that they were
not being paired with another person at a different site. Immediately after subjects completed the experiment, they were asked about it. One question asked whether their counterpart was another person or a computer: 62.4 percent answered they were playing against another individual. While this rate is low, analysis not reported here indicates that this made no difference over the distribution of choices. Part of this may stem from the fact that all subjects were paid for their choices and as a consequence they were attuned to the decision problem.

Compared with the single-shot MRS game, again our subjects are less likely to choose the cooperative left branch as their first choice. They do so only 44.4 percent of the time (see Table 2). By comparison, when facial images are introduced, there is some impact on subject's first choice. The image with a smile results in a 10 percent decrease in the choice of left. The neutral image results in a 24 percent increase in choosing left. One should be cautious interpreting this latter finding because of the small number of subjects. Nonetheless there is a large increase in the percentage choosing that branch.

As with the pilot studies, subjects were asked, after the fact, how they rated their counterpart in the experiment. The second and third rows show that when subjects were given an image they rated their counterpart differently. The largest differences, however, were between the image with a smile and images with either a neutral or angry expression. There was no important difference between the latter two images -- unlike the ratings given in the pilot experiment with the identical images.

These findings are quite puzzling. While there is some consistency in the ratings of counterpart images when compared with the pilot study, there is no consistency in the behavioral response ordering. The percentage of subjects choosing left is substantially lower when they view a positive (smile) expression than when observing a neutral image. It is clear that by exposing subjects to an image behavior is changed. However, these computer controlled experiments do not point to any consistency in what the social signal means.

Discussion.

Both the pilot and computer controlled experiments support the conjecture that images produce a reputational prior that is different from that when there is no information about the counterpart in the experiment. What is not clear is how that reputational prior is processed by subjects and how it is systematically related to choices.

There are several features that are striking in these experiments. First, we consistently find that more than a third of the subjects reported they were confused by the
instructions. Although we have implemented several checks for comprehension in the experiments, we feel they are insufficient and we are skeptical whether subjects clearly understood the game.

Second, the game form used in both experiments is sufficiently nuanced that if subjects do not understand the game, it is unclear how to interpret outcomes. Following the pilot experiments, subjects were asked to write a sentence or two indicating the basis for their choice. A disconcertingly large number of subjects indicated the choice was random. The results clearly show that this is not the case. Nonetheless, if subjects failed to comprehend the game because the instructions were unclear, then a great deal of noise is introduced into these experiments.

Finally, it is clear that these experiments have not tested whether cooperative or competitive mental modules are activated. While we anticipated that particular images would signal distinct reputational priors to subjects, it is not clear that the images selected do what we expected. Additional analysis of the images (which is forthcoming) remains. At best we can conclude that some form of a reputational signal is important. At this point we are not certain about how those signals affect behavior. We suspect that there is a great deal of noise in these data. That noise is generated by unclear instructions, incomplete understanding by subjects about the nature of the game, and uncertainty about what the stimuli are activating.
Bibliography


### Table 1
Pilot Experiment

<table>
<thead>
<tr>
<th>Percentage Left</th>
<th>No Face</th>
<th>Smile</th>
<th>Neutral</th>
<th>Angry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39.5</td>
<td>44.7</td>
<td>27.6</td>
<td>28.2</td>
</tr>
<tr>
<td>Average Rating Happy/Sad</td>
<td>1.63</td>
<td>1.39</td>
<td>2.02</td>
<td>2.39</td>
</tr>
<tr>
<td>Average Rating Cooperative/Uncooperative</td>
<td>1.41</td>
<td>1.53</td>
<td>2.05</td>
<td>2.23</td>
</tr>
<tr>
<td>Total Number of Subjects</td>
<td>76</td>
<td>103</td>
<td>116</td>
<td>78</td>
</tr>
</tbody>
</table>

### Table 2
Computer Controlled Experiment

<table>
<thead>
<tr>
<th>Percentage Left</th>
<th>No Face</th>
<th>Smile</th>
<th>Neutral</th>
<th>Angry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44.4</td>
<td>34.4</td>
<td>68.4</td>
<td>41.9</td>
</tr>
<tr>
<td>Average Rating Happy/Sad</td>
<td>1.74</td>
<td>1.22</td>
<td>1.89</td>
<td>1.94</td>
</tr>
<tr>
<td>Average Rating Cooperative/Uncooperative</td>
<td>1.63</td>
<td>1.43</td>
<td>1.78</td>
<td>1.77</td>
</tr>
<tr>
<td>Total Number of Subjects</td>
<td>27</td>
<td>32</td>
<td>19</td>
<td>31</td>
</tr>
</tbody>
</table>
Figure 1
Extensive Form Game Used in Experiments
Figure 2
Images Used as Stimuli in Experiment

Smile  Neutral  Anger