Judging a Book by its Cover: Beauty and Expectations in the Trust Game

Rick K. Wilson and Catherine C. Eckel

Political Research Quarterly 2006; 59; 189
DOI: 10.1177/106591290605900202

Judging a Book by its Cover: Beauty and Expectations in the Trust Game

RICK K. WILSON, Rice University
CATHERINE C. ECKEL, University of Texas, Dallas

This research examines one mechanism by which people decide whether to trust strangers. Using a laboratory setting that provides subjects with controlled information about their counterparts, we test whether attractive subjects gain a "beauty premium" in a game involving trust and reciprocity. Attractive trustees are viewed as more trustworthy; they are trusted at higher rates and as a consequence earn more in the first stage of the game. Attractiveness does not guarantee higher earnings, as we find a "beauty penalty" attached to attractive trusters in the second stage of the game. This penalty arises because attractive trusters do not live up to expectations of them on the part of the trustees. Trustees withhold repayment when their expectations are dashed. `This punishment is larger when the disappointing truster is attractive.

Our mothers taught us "beware of strangers" and "don't judge a book by its cover." The first maxim was intended to protect us from being put at risk by an unknown person. The second warned us not to be fooled by appearances, but to look deeper into a person. Both have important implications for a decision whether or not to trust another person. Trust involves putting one's well-being into the hands of another. Our mothers' advice warns us to be wary of those we do not know and to scrutinize carefully a person before trusting them. If taken at face value we should only trust those with whom we have had first-hand experience and those whose reputation is spotless. Such a cautious approach is likely to rule out many potentially profitable opportunities, however.

Trust is a key concept in political science. Fenno (1978) and Bianco (1994) argue that it is central for defining the relationship between representatives and the represented. Ostrom (1990, 1998) points out that it is crucial for solving many common property resource dilemmas. Among strangers, trust is considered to be the glue that holds together social relations (see for example Putnam (2000), Hardin (2002), among many others—although these scholars would not agree about the mechanism by which trust works its sticky magic).

The research reported here examines snap judgments—one mechanism by which people decide whether to trust strangers. Subjects in a laboratory experiment prove to be willing to judge a book by its cover, trusting more attractive strangers even though, from a strategic standpoint, they should not. This study uses a setting that provides subjects with controlled information about their counterparts. We observe the effect of attractiveness on the willingness to trust and to reciprocate trust in a bargaining game with financial stakes. We find that attractive subjects gain a "beauty premium" in that they are trusted at higher rates (and as a consequence earn more). Attractiveness, however, does not guarantee higher earnings as we find a "beauty penalty" when trust is reciprocated.

NOTE: This article was delivered as a paper at the Western Political Science Association Meetings, March 27-29, 2003, Denver. Thanks go to Selhan Garip, Sezi Anac, Bing Yaun, and Johanna Dunnaway for help in conducting the experiments. Lauren Porzelt from the Russell Sage Foundation and Christy Aroopala from Rice University provided help with the manuscript. Support for this research comes from grants by the National Science Foundation (SES 0318110 and SES 0318116) and the Russell Sage Foundation.

Political Research Quarterly, Vol. 59, No. 2 (June 2006): pp. 189-202

Motivation

Trust involves a decision that makes one vulnerable to the actions of another. In order to know if it is wise to trust another, it is important to assess whether that counterpart is trustworthy. An accurate assessment of trustworthiness, however, depends on knowing whether the counterpart has been trustworthy in the past and believing that the counterpart will continue to be trustworthy. In most societies trust is supported by institutions and social arrangements that monitor the actions of others and that minimize the cost of information about what others have done in the past. For example, tort law and gossip both can be effective for sustaining trust. In other contexts, in which there are ongoing relationships among actors, trust can be fostered and sustained through reputation and experience. Trust, too, can be sustained among neighbors where there is a shared history of beneficial exchange. All of these mechanisms for facilitating and enhancing trust have dominated the literature on social capital. See Granovetter (1985), Coleman (1988), and Putnam (2000).

What happens when people are strangers? Under such a circumstance there is no reputational information about the counterpart. It is difficult to form any expectations based on experience or social connection. When dealing with strangers it would seem prudent to follow the age-old adage of "beware of strangers." Yet people often trust those they do not know and they do so, on a belief that they can "read" the trustworthiness of others based on something signaled by their counterpart.

A Trust Game

Considerable research has focused on measuring individual attitudes toward inter-personal trust, generalized trust and institutional trust [see the survey by Levi and Stoker (2000) or books by Hardin (2002) and Uslaner (2002)]. This work shows that trust as an attitude is widespread, although its long-term stability and behavioral consequences are unclear. Another research stream has focused on experimental methods, using a two-person sequential game to distill components of trust (see the recent edited volume by Ostrom and Walker 2003). First published as the "investment game" by Berg, Dickhaut, and McCabe (1995), the experiment provides an endowment for Player A who decides how much of the endowment (if any) to pass to Player B. Whatever is sent is tripled in value by the experimenter. Player B then decides how much of the tripled amount (if anything) to return to Player A. Once the decisions are made, the experiment is complete and both players take their money and go home. For myopic maximizers the prediction for the game is clear: the unique subgame perfect equilibrium is for Player A to never send anything to Player B. If anything is sent Player B will keep it all. There is no subsequent move that Player B needs to fear, so there is no penalty for acting self-interestedly. Using backward induction, Player A understands this and should send nothing. Of course, this game introduces an important tension in that both players could be better off if Player A invested and Player B reciprocated by returning more than what is invested. This particular game is thought to measure concepts of trust and trustworthiness. Any amount passed to player B is viewed as a measure of trust on the part of player A. Anything returned to Player A is viewed as trustworthiness on the part of Player B.[1]

The game is very simple, the optimal strategies are transparent and experiments using this design show that the old adage cautioning against trusting strangers is ignored. Camerer (2003), when surveying the experimental literature, finds that typically about half of Player A's endowment is invested. As well, there is trustworthiness, with Player Bs returning something to their (anonymous) counterpart. Trust, however, does not pay on average, with a return that often ranges from 80 to 95 percent of what was invested. These findings are not unique to student subjects in the United States. For example, Yamagishi, Cook, and Watabe (1998) find a good deal of trust among both Japanese and U.S. subjects [see also Hayashi et al. (1999). Bahry and Wilson (2004) depart from the usual sample of college students and conduct trust experiments on population samples in the Russian Republics of Tatarstan and Sakha (Siberia). They find substantial levels of trust and trustworthiness and those patterns resemble findings from the United States.[2] Motivated by the possibility that subjects might respond to a group of familiars (non-strangers), Eckel and Wilson (2004) designed an experiment that guaranteed subjects did not know one another. Conducting the experiment via the internet across states, they too find considerable trust and trustworthiness. The general conclusion from these studies is that people are willing to trust, there is substantial reciprocated trust and there is plenty of heterogeneity in the decision to trust. Put simply, people trust strangers.

This raises the question of whether people also ignore the second piece of advice: not to judge a book by its cover. Several experimentalists have moved away from strictly anonymous pairings of subjects in order to provide limited information about subjects' counterparts. This is to answer the question of whether trusters infer trustworthy attributes about their counterpart based on "cheap talk" signals. For example Scharlemann et al. (2001) and Eckel and Wilson (2003b) find that something as simple as a smile invites trust. Croson and Buchan (1999) find that women reciprocate more than men in trust games but that there is no difference between the two in the decision to trust. Chaudhuri and Gangadharan (2002) obtain the same finding with respect to reciprocity, but they find that men are more likely to trust than women. Neither experiment gives subjects information about the gender of their counterpart, so it is unlikely that subjects are making judgments about their counterpart based on what they know or see. On the other hand, Scharlemann et al. (2001) have subjects observe the gender of their counterpart and they find that when males are paired with females trust and trustworthiness increase. Eckel and Wilson (2003a) have subjects view the photo of their counterpart and find that Caucasians do not trust African-Americans, largely because Caucasians expect them to be less generous. Likewise, African-Americans who trust are reciprocated less, pointing to underlying patterns of discrimination.[3] It appears that subjects use the information that is available to them and draw inferences about their counterpart.

Judging a book by its cover, however, seems a risky strategy. Relying on deliberate signals like smiles or the signaling properties of characteristics such as gender or race and ethnicity seems naive. In the language of game theorists such signals are little more than "cheap talk" in that they convey no credible information about the strategy of the counterpart. A smile is easily mimicked but perhaps not easily

[1] Some readers may be skeptical about whether this experiment measures trust. However, it has all of the properties of a trust decision: individuals make themselves vulnerable to another. Likewise the experiment taps trustworthiness on the part of Player B. This experimental design has proven to be a very rich source for models concerned with trust and for fleshing out a notoriously slippery concept. The skeptical reader is directed to Ostrom and Walker (2003) or Camerer (2003).

[2] Among other cross national studies producing similar findings are those by Koford (2001) (Bulgaria), Fershtman and Gneezy (2001) (Israel), Buchan, Croson, and Dawes (2002) (China, Japan, Korea, and the U.S.), and Karlan (2003) (Peru).

[3] Similar findings also appear between Blacks and Whites in South Africa (Burns 2003) and across different ethnic groups in Israel (Fershtman and Gneezy 2001).

faked—see Ekman (2003)—and as a consequence should serve both those who are and those who are not trustworthy. Gender and race are more difficult to mimic, but may be no more informative than a smile. Although people may have powerful stereotypes (expectations) about populations with these phenotypic markers, they may do little to inform the decision about trusting a specific individual. We are interested in whether people infer something about their counterpart based solely on a surface judgment, and whether these judgments are correct. To this end we turn to an easily observable, seemingly non-credible, but difficult to mimic, aspect of people: their attractiveness.

Why Trust Atractiveness?

Does attractiveness signal trustworthiness? There is ample research suggesting that people attribute a variety of characteristics to others on the basis of physical attractiveness. See the extensive reviews by Eagly et al. (1991), Feingold (1992) and Langlois et al. (2000). In general these studies suggest that people attribute positive characteristics—intelligence, competence, leadership skills, etc.—to attractive persons. In economics, multiple studies show that attractiveness pays off in the market place, and this has been tagged a "beauty premium." See Hamermesh and Biddle (1994); Biddle and Hamermesh (1998). Political scientists, too, have noted a beauty premium for candidates. Dating at least from the famous televised "dark shadow" debate between Kennedy and Nixon, political scientists have been intrigued by the possibility that looks can drive election results (at least at the margin). A survey of the literature by Ottati and Deiger (2002: 79) leads them to conclude:

> current analysis suggests that physical attractiveness activates a stereotype that influences judgments of a candidate in a relatively direct manner. However, it is possible that the effect of the physical attractiveness stereotype on candidate evaluation is mediated by selective encoding, biased interpretation, biased elaboration, or selective retrieval of a candidate's issue stances.

As a collection, these studies point to middle level theories about how stereotypes might mediate behavior and judgments. These theories point to ways in which people develop stereotypes about attractiveness. This gives rise to the idea that there is a beauty premium for attractive people, based on expectations about what attractive people are like (e.g., they have greater leadership skills, they are friendlier, they are more competent, etc.). However, few scholars have addressed the deeper question of why attractiveness ought to be something that humans pay attention to and why it is so resilient. Exceptions to this are found in evolutionary psychology. Drawing on animal behavior and inferences based on selection pressures in evolution, three types of arguments are offered as to why attractiveness might be something that catches the human eye. These arguments focus on good genes, parental investment and status, and suggest reasons why people are pre-disposed to attend to beauty.[4]

The good genes argument points to attractiveness as a signal of genetic quality for purposes of mating. The parental investment argument points to differential investment strategies by parents in their offspring. Early choices about nutrition lead to life-long effects on characteristics like height and symmetry—features that others find to be attractive. Finally, there is an argument that attractiveness is a visible marker conferring status. All three arguments similarly focus on attractiveness as a visible signal for humans. It is a signal to which we are predisposed to be attentive (often for purposes of mating and genetic fitness). In the conclusion we return to these arguments and what they mean with respect to our findings.

Can attractiveness signal trust or trustworthiness? In a game theoretic sense attractiveness is a credible signal only if it contains information about whether a person can be trusted, and if it cannot be faked. As the discussion above suggests, people are attentive to attractiveness and they may well infer that such people are trustworthy. It is well established that people ascribe positive characteristics to those who are more attractive (Feingold (1992); Langlois et al. 2000). If attractive people act in a manner consistent with expectations, then attractiveness is a credible signal.

Do strangers use attractiveness as means of discriminating their trust? If so then do attractive second movers obtain a beauty premium by being trusted more than unattractive second movers. Are attractive trustees more likely to reciprocate? Does reciprocity depend on the attractiveness of the truster? We examine these questions in a laboratory trust game experiment.

RESEARCH DESIGN

There are two components to this research. In the first component subjects are photographed, they participate in a series of trust games, their attitudes toward risk are tested with a variety of instruments and they answer questions from an attitudinal survey eliciting individual characteristics and demographics. In the second part of the research the photographs of subjects in the trust game are evaluated by an independent group of subjects. These evaluations include assessments of the attractiveness of the photographed subject.

The behavior of subjects was tested using a trust game experiment. The experiment included three main components: a trust game, an assessment of risk attitudes, and responses to a survey of attitudes toward trustworthiness and altruism. A total of 206 subjects, half from Virginia Tech, 42.2 percent from Rice University and 7.8 percent from North Carolina A & T, participated in eight experimental sessions. Subjects were recruited from introductory

[4] See the general discussion by Langlois, et al. (2000). Berry (2000) offers an excellent overview of the issues discussed here. For empirical work, see Zebrowitz and Rhodes (2004).

classes in Principles of Economics at Virginia Tech and NCAT and from dining halls at Rice University. They were told to report at a specific time to a laboratory at their respective locations. The number of subjects in a session ranged from 10 to 32. Subjects were 55.8 percent male and just under 94 percent of subjects were between the age of 18 and 22. Care was taken to recruit an ethnically diverse subject pool.[5] The bulk of the subjects (62.6 percent) were Caucasian, 15.0 percent were African-American, 12.6 percent Asian-American, 5.3 percent Hispanic and the remaining 4.4 percent self-identified foreign nationals.

When subjects arrived at the lab they were asked to sign a consent form and given a card assigning them to a specific computer. Subjects then posed for four pictures—two neutral and two smiling expressions. Oral instructions were read to subjects at the outset and then subjects went through a set of self-paced, computerized instructions. In a post-experiment questionnaire a little over 90 percent very strongly or strongly agreed that the instructions were clear. Once subjects began the experiment, no talking was allowed. Subjects were asked to raise their hands if they had a question or problem, and the experimenter would answer their questions privately.

Prior to beginning the trust game subjects chose one of their four pictures. They were told their photograph would be seen by others during the course of the experiment. Subjects were not told what the experiment entailed when they made their choice. Once the photograph was chosen, subjects viewed group photographs designed to ensure beliefs that they were facing a real on-line counterpart.[6] Subjects completed an on-line survey designed to measure attitudes toward risk, the Zuckerman Sensation-Seeking Scale (SSS), form V Zuckerman (1994). Subjects earned ten experimental lab dollars for completing the survey (the exchange rate was two lab dollars for each U.S. dollar).

Once the trust experiment began subjects were told that they would be paired with up to ten different people.[7] They were also told that only a single counterpart was guaranteed to be on line with them and it was that decision period for which the subject would be paid. Subjects were told that their "real" counterpart would not be revealed until the end of the trust game. Before they were informed, the subjects were asked to guess their counterpart and if they guessed correctly they would earn $1.00. First movers only managed this 11.5 percent of the time and second movers did so 12.4 percent of the time. The two on-line players were always paired in the first period, although the subjects did not know this. The "additional" photographs were from subjects who participated in earlier sessions of the experiment or from images used in a similar experiment. For details on that experiment, see Eckel and Wilson (2004). All of the subjects were photographed in the same way to ensure consistency in display of the manipulation.

Subjects were always assigned the same position in the experiment and they were told this. As with the trust game developed by Berg, Dickhaut, and McCabe (1995), first movers (trusters) could send any whole dollar amount of their survey earnings to their counterpart. The experimenter tripled the amount sent. Prior to making their decision the truster viewed the photograph of the second mover (the trustee). We took pains to ensure that the subjects were strangers and that they were unlikely to interact in the future. We achieved this by matching subjects across different Universities (usually separated by a time zone). In order to minimize learning and adjusting expectations, these trusters received no feedback between decisions (i.e., first movers learned nothing about what was sent back).

The trustees also viewed the photograph of their counterpart. Prior to being told how much the truster sent, trustees predicted how much would be sent (and they were paid $1.00 for an accurate forecast). The trustee then decided how much, if any, of the tripled amount would be returned. Trustees were only paid for a single pairing of their on-line counterpart.

Once the trust game decisions were completed subjects finished the risk-assessment element of the experiment by making a series of risky decisions. The risky decision task replicates the risk instrument designed by Holt and Laury (2002) in a computerized environment. Subjects face a series of paired lottery choices with one of the decisions chosen at random and played out for payment. Although several risk instruments were used in this experiment, we do not report the results here (see Eckel and Wilson 2004). In the last component of the experiment subjects responded to a questionnaire collecting demographic data and measurements of their attitudes toward trust and altruism. Once finished subjects were paid, one at a time, in private and sent on their way. Overall, subjects earned $15.10 US for approximately one hour in the laboratory.

Photograph Evaluations

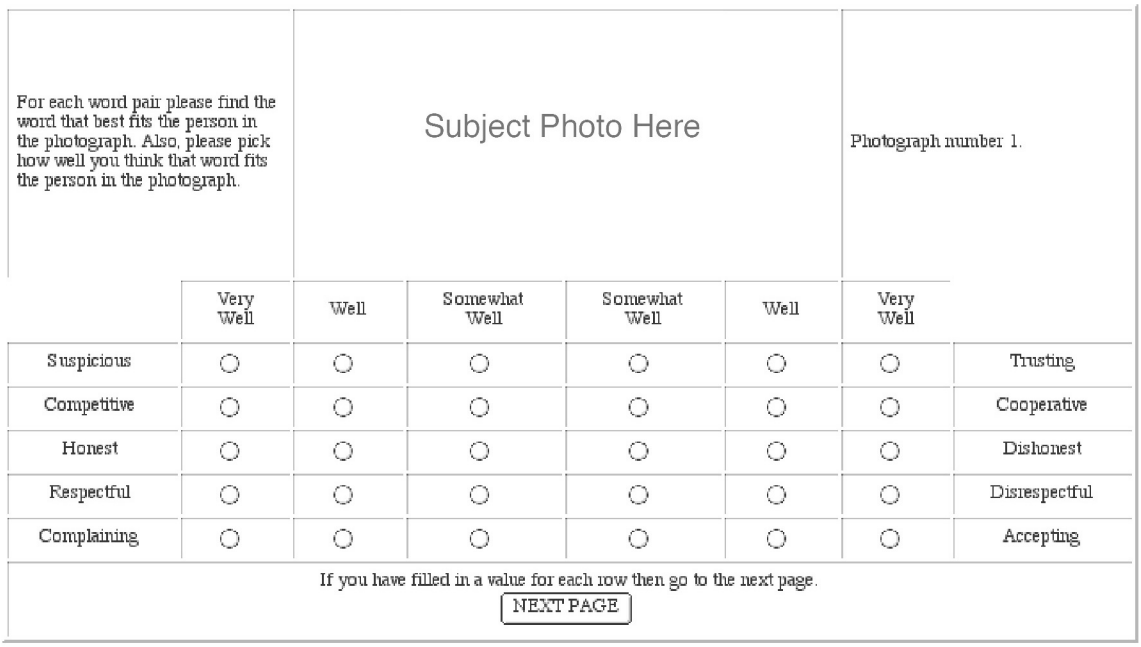
Following the trust experiment, an independent sample of subjects evaluated the photographs used in the trust game.

[5] One session was conducted between Virginia Tech and North Carolina A & T. The latter was chosen because it is an Historically Black University with a heavy emphasis on engineering. This provided a useful match with subjects from Virginia Tech.

[6] In this portion of the experiment a subject was randomly selected to name a code word that was sent to the other site by the experimenter. The other site's codeword was written on a large sheet of paper and a group photo was taken. The codeword and subjects were visible (although not the latter's faces) and that photograph was uploaded to the server. Subjects then viewed the images at both sites, with their own codeword displayed at the other location. As shown in Eckel and Wilson (2006) this ensures that subjects believe they are facing an on-line counterpart. A post-experiment questionnaire item indicated that 85 percent of the subjects believed their on-line counterpart was real.

[7] In the first session subjects participated in six trust games; in the second session subjects participated in eight trust games; in the remaining sessions subjects participated in ten trust games. At first we were concerned with how much time it would take to run the additional trust games. A pilot test indicated it might take a considerable amount of time. However, once we began running the experiment it was clear that subjects could easily go through ten different trust games.

FIGURE 1
A SAMPLE SCREEN AND SUBSET OF THE WORD PAIRS USED IN THE EVALUATION EXPERIMENT



The evaluators had not participated in the trust experiment and were only asked to evaluate a set of images. A total of 296 subjects participated: 56.4 percent were male. Subjects were recruited over the internet and from large classrooms at different sites. Each subject was asked to rate between 15 and 24 photos on a 15 word-pair items scale and was paid between $.25 and $.50 per photo.[8] There were 230 photos with a total of 5,216 evaluations. Photos and their order were randomly assigned to each subject. Subjects spent an average of 80.2 seconds per photo (with a standard deviation of 64.5 seconds). In order to mimic the trust experiment, subjects were recruited from three different universities. Raters only evaluated photographs from a different site. Because photos were randomly assigned and because responses at each site varied, the number of raters per photo ranged from 9 to 42, with an interquartile range between 18 and 27.

Each image was evaluated using 15 opposite word pairs. A sample of five word-pairs and the screen used by subjects is shown in Figure 1. The left/right order of the word pairs was randomly fixed prior to the experiment. The order of presentation of the word pairs was randomized across each photograph for each subject, providing control over response-set bias in the word pairs.

We focus on the attractiveness rating. One of the word-pairs was "unattractive-attractive" and like all of the words was evaluated on a six-point scale, with the word fitting the person in the photograph either "Very Well," "Well" or "Somewhat Well."

[8] Subjects were paid different amounts for the evaluations depending on whether they participated in additional experiments. We find no difference in evaluations between those who were paid more or less.

RESULTS

We begin with descriptive statistics concerning the overall levels of trust. Subjects sent an average of $4.91 in Lab dollars (compared with the $5.16 sent by subjects in the BDM experiment). Table 1 provides the mean and standard deviation for these data. While subjects send slightly more to their on-line counterpart than to off-line players the difference is not significant ($t = .267$, $p = .79$, $df = 954$); therefore these data are pooled. Unlike the findings by BDM, in the aggregate, trust pays. Subjects returned $6.41 on average, while for BDM an average of $4.66 was returned.[9] Although somewhat more is returned to robots than to an on-line counterpart, the difference is not statistically significant ($t = .581$, $p = .56$, $df = 900$). Of course, these are only averages and as is clear from the standard deviations there is considerable heterogeneity in what individuals both send and return.

Figure 2 below produces a scatterplot of what was sent and the amount returned for both on-line counterparts. The circles represent what was sent and what was returned. Because only whole dollars could be sent and returned many of these points lie on top of each other. They have been perturbed by a small value in order to give a sense of

[9] This result is consistent with results reported by Burnham (2003) in which photographed images were used in a "dictator game." In that experimental design subjects who viewed images of their counterpart sent more than those who did not see a counterpart. Burnham's design has no expectation of reciprocity, but it clearly points out that the use of photographed images increases the amount of money given to another.

TABLE 1
AVERAGE AMOUNT IN LABORATORY DOLLARS SENT BY FIRST MOVERS AND THE AVERAGE AMOUNT RETURNED BY SECOND MOVERS (STANDARD DEVIATION IN PARENTHESES)

	Overall Average	Average for "Real" Counterpart	Average for "Other" Counterparts
Average Amount Sent to Counterpart	$4.93 (3.24) n = 943	$4.99 (3.06) n = 103	$4.89 (3.26) n = 840
Average Amount Returned	$6.41 (5.90) n = 888	$6.10 (5.17) n = 98	$6.44 (5.98) n = 790

The numbers of the first and second movers vary because second movers who received nothing are excluded from these calculations.

the distribution. The solid line on the figure indicates a "break even" point at which the truster was returned exactly what was sent. Circles above the solid line meant that trust paid off.

Figure 3 displays similar data only this time from the standpoint of second movers who responded to what was sent by either on-line counterparts or robots. Again the circles constitute an amount sent and what was returned, and the solid line represents an amount returned that was equal to what was sent. Both of these figures suggest that there is substantial heterogeneity among subjects.

Not only is there heterogeneity among subjects, but also within subjects. One feature of this design is that it allows us to examine the behavior of individuals across many decisions. We find that 25 of 103 first movers (24.3 percent) never change their strategy during the course of the experiment, always doing the same thing. Almost half always sent their full endowment (12 of 25), two subjects never sent anything and five always sent half of their endowment. The other first movers all have some variation in what they chose to send.

There is greater heterogeneity among the second movers. Only 7 of the 103 reciprocators kept the same strategy for returning (5 always returned nothing, 1 returned exactly what was sent and 1 returned everything). These data indicate that subjects are discriminating among their counterparts. On what basis are people doing so?

Is There a Beauty Premium?

Attractiveness, as we have argued above, is a fundamental part of the human calculus and it may explain part of the heterogeneity observed in the trust data. We calculate an attractiveness rating of the photographs derived from the independent sample of evaluators. A transformed measure is used because there are two problems associated with using a direct measure of attractiveness. First there are individual differences in evaluations. For example, as noted in the meta-analysis by Jackson, Hunter, and Hodge (1995), females tend to give higher evaluations than males. Indeed, we find this is the case: for example females give significantly higher ratings than males ($t = 8.27$, $df = 5214$, $p < .001$). The second problem noted in the literature is that, irrespective of the sex of the evaluator, images of females are rated higher than those of males and this we also find ($t = 13.87$, $df = 5214$, $p < .001$). To deal with the first difference we mean-center each rater's attractiveness ratings. This has the effect of accounting for individual differences, with ratings set to each individual's average score. The rating was calculated by:

$$\text{Attractiveness}_i = (x_{ij} - \bar{x}_j) \text{ where } \begin{cases} i = \text{ith photo} \\ j = \text{jth evaluator.} \end{cases}$$

This transformation does not get rid of the problem that photos of females are rated higher than those for males, regardless of the sex of the evaluator. To control for this, the mean-centered attractiveness ratings were normalized for the sex of the photographed image. This meant taking the individually corrected ratings and then accounting for variation across the gender of the photograph. Z-scores were generated for each image, accounting separately for the means and standard deviations of female and male images. (Later, we return to multivariate analysis of ratings unadjusted by gender.)

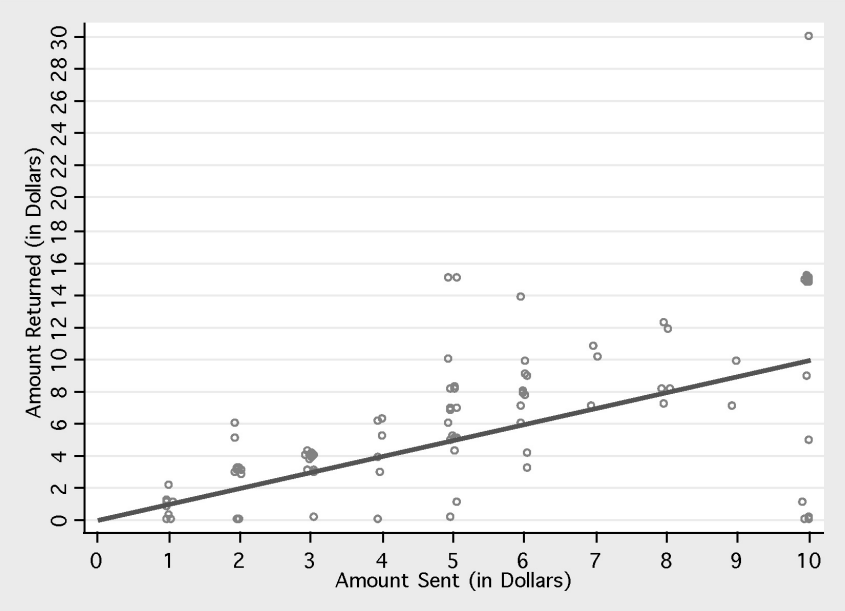
Consistent with the meta-analysis by Langlois et al. (2000) we find that subjects are in substantial agreement concerning their evaluations. Using Cronbach's alpha as a measure of consistency we obtain values of .80 for the original values for attractiveness and .82 for the normalized values.[10] Once these normalized scores are calculated we find no differences across the sex of the evaluator nor the sex of the photographed image.[11]

Are more attractive people trusted at higher rates than less attractive people? We begin by looking at the amount sent in

[10] Because there were different numbers of evaluators for each photo, a bootstrap method was used to calculate the values of Cronbach's alpha. Sample evaluations of 9 subjects were used for each photo. A total of 50 runs were used to compute the average scores.

[11] Student's t-tests of differences of means of the normalized attractiveness scores are as follows. For the sex of the evaluator, ($t = .327$, $df = 5214$, $p = .74$) and for the sex of the photograph ($t = .00$, $df = 5214$, $p = 1.00$).

FIGURE 2
AMOUNT SENT AND PERCENTAGE RETURNED FOR ON-LINE SUBJECTS ONLY



Circles represent a pair of subjects, with the amount sent on the horizontal axis and the amount returned on the vertical axis. Circles on or above the line indicate that trust paid off.

the trust game to the most and least attractive.[12] Trust is measured as the amount sent. After all one should send money only if one is confident the counterpart is trustworthy.

The first column of Table 2 compares the amount sent to those who are the least and most attractive. These two groups constitute those who are rated as one standard deviation above or below the mean for the attractiveness ratings. On average, more is sent to attractive than unattractive subjects. There is an average difference of $.34 reflecting a beauty premium of 7.3 percent. Using a one-tailed test (we expect more attractive to be treated differently) we find modest support for a beauty premium.

Likewise it appears that trusting attractive counterparts is justified. It is important to keep in mind that the amount returned is related to the amount sent. For purposes of comparison we focus on the percent of the total amount returned. Because whatever was sent was tripled, a percentage greater than 33.3 indicates more was reciprocated than sent. Column two of Table 2 shows that more attractive trustees reciprocate more. Apparently, trusting attractive subjects pays off.

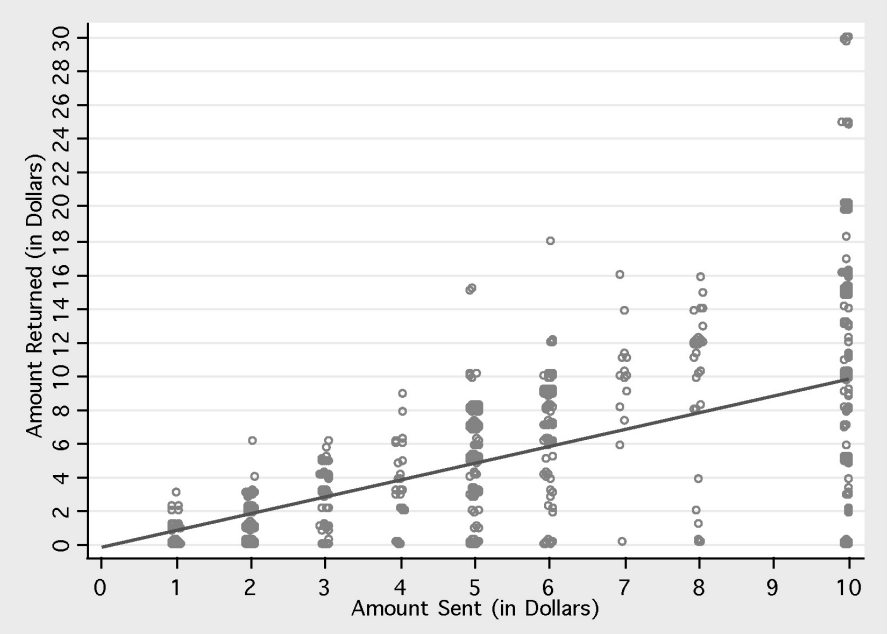
If the analysis ended at this point we would conclude that there is a pervasive beauty premium. However, so far we have only looked at the characteristics of the trustee. Is it the case that attractive trusters get more back? The first column of Table 3 compares the average percentage returned to the most and least attractive first movers. Again, the two groups are split using one standard deviation above and below the mean using the normalized score. Here the findings are contrary to expectations. Unattractive subjects receive an average of 35 percent of what was sent, whereas for attractive first-movers, trust does not pay, with around 30 percent returned. This indicates that there is a beauty penalty and not a beauty premium at work.

The origin for such a penalty may be tied up in the expectations about the first mover. Prior to seeing what was sent, trustees were asked to predict how much was going to be sent and this guess is used to tap expectations. The second column of Table 3 compares these expectations for subjects broken out by the attractiveness of the truster. It is clear trustees have higher expectations for their more attractive counterparts.

It seems reasonable to expect that dashed expectations will be punished. If so, this may help explain the "beauty penalty" results for reciprocation. The simplest way to measure this is to look at instances in which expectations were dashed and when they were exceeded (we ignore the 13.4 percent of the cases when expectations were met). If this measure is cross-tabulated with the least and most attractive, the source of dashed expectations becomes clear. Among those with dashed expectations, only 27.4 percent faced a less attractive first mover and 72.6 percent faced an

[12] It is common in the attractiveness literature to look at the extremes. Often the upper and lower quartile is used. We use those photos for which the composite evaluation is one standard deviation above or below the mean. This composite measure is the average attractiveness rating derived from the independent sample of evaluators. The normalized value is used which controls for the gender of the photograph and individual characteristics of the evaluator. These ratings are not related to the race or ethnicity of the photograph, although African-American images tend to be rated slightly more attractive than others (although this difference is not statistically significant).

FIGURE 3
PERCENTAGE RETURNED BY HUMAN SECOND MOVERS



Circles represent a pair of subjects, with the amount sent on the horizontal axis and the amount returned on the vertical axis. Circles on or above the line indicate that trust paid off.

attractive first mover. For those whose expectations were exceeded the differences are not so stark with 56.9 percent facing a less attractive and 43.1 percent facing a more attractive counterpart.[13]

Figure 4 shows that the relationship between attractiveness and dashed expectations is not due to just these main effects. The figure plots the average return given dashed and exceeded expectations. At the same time this is broken out by trusters who are one standard deviation above or below the mean for attractiveness. It can be seen from the figure that attractive counterparts are penalized more than unattractive counterparts when expectations are dashed (returning 25.8 percent to the former and 34.9 percent to the latter). By contrast there is little difference between attractive and unattractive counterpart when expectations are exceeded (returning 35.2 percent and 33.3 percent, respectively). There is clearly an interaction with attractive counterparts and expectations, while the same is not true for less attractive trusters.

In sum these results point to a beauty premium for trustees and a beauty penalty for trusters. In the trust decision, more attractive second movers receive a 7.3 percent "bonus" compared to their less attractive colleagues. In addition, it appears that attractiveness acts as a signal of trustworthiness, as more attractive trustees reciprocate at higher rates. Simply put, they receive more and they return more. Yet first movers suffer a beauty penalty. This is because second movers expect attractive trusters to send more than they do. Dashed expectations are punished, but only for those who are attractive.

Multivariate Models

Parsing the data into extremes of attractive and unattractive subjects may overstate the results. As well, relying only on an "attractiveness" measure may miss other aspects of the photographs that affect trust. For example, trust and trustworthiness may be related to whether the subject was smiling, the jewelry that was worn or even the race and ethnicity of the subject. To test for this possibility a series of regressions are presented. The dependent variable for the first model is the amount that a subject sent to the counterpart. Because individuals made a series of decisions (and these decisions are not independent) we use a random effects model.[14] The random effects are estimated for each individual decision maker.

Because we control directly in the regression for gender-related differences, the attractiveness variable that we use is uncorrected for gender. We use the mean-centered measures, which are adjusted only for differences in averages across

[13] A simple chi-square test points to the significance of the relationship with $\chi^2(1) = 33.76$, $p < .001$.

[14] A random effects regression treats the data as a panel, controlling for individual effects. Because we expect characteristics of the decision maker to interact with characteristics of the counterpart we do not use a fixed effects model

TABLE 2
AVERAGE AMOUNTS IN LABORATORY DOLLARS SENT BY TRUSTERS, CONTINGENT ON THE ATTRACTIVENESS RATING OF THE TRUSTEE (STANDARD DEVIATION IN PARENTHESIS).

	Amount Sent to Trustee	Percent Returned by Trustee
Unattractive Counterpart	$4.64	32.77
(One Std. Dev.)	(3.18)	(21.56)
Attractive Counterpart	$4.98	35.99
(One Std. Dev.)	(3.24)	(19.66)
t-test	t = 1.22, df = 581, p < .11	t = 1.76, df= 523, p < .04

TABLE 3
PERCENTAGE OF THE INVESTMENT RETURNED TO AND THE AMOUNT IN LABORATORY DOLLARS EXPECTED FROM ATTRACTIVE AND UNATTRACTIVE TRUSTERS (STANDARD DEVIATION IN PARENTHESIS).

	Percent Returned to Truster	Amount Expected by Trustee*
Unattractive 1st Mover	35.25	$5.15
(One Std. Dev.)	(21.67)	(2.87)
Attractive 1st Mover	30.38	$5.50
(One Std. Dev.)	(23.71)	(2.96)
t-test	t = –2.55, df = 594, p = .995	t = 1.45, df = 621, p = .07

Note: *This column includes second movers who received nothing from the first mover.

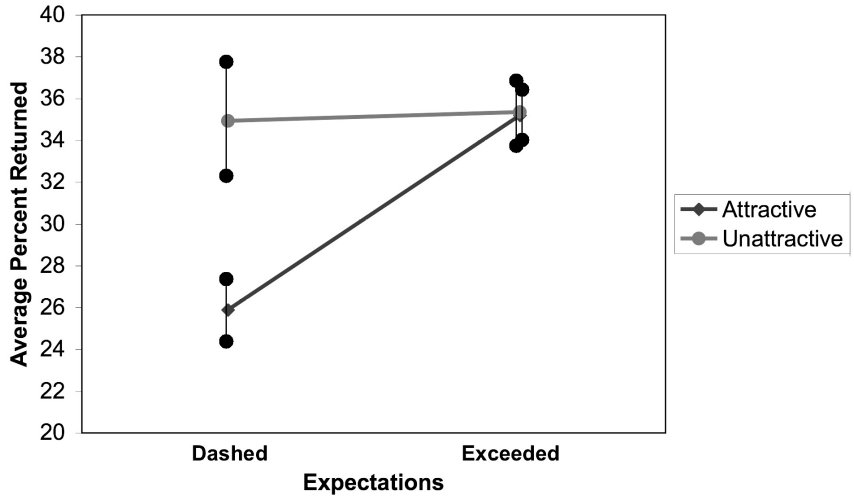
evaluators, and not for the differences in ratings of male and female images. The mean-centered scores are averaged across evaluators for each image as before, and this score is used as the attractiveness rating. Table 4 provides descriptive statistics for this and other independent variables.

In the first model we are concerned with the gender pairings of the subjects. Different pairings of males and females may have large consequences for strategic calculation. For example, males may respond differently to male than female counterparts. To control for this possibility we create variables for each gender pairing and player position (the male truster/male trustee pairing is the omitted category). Each pairing is coded as a dummy variable. It may also be the case that having an African-American, Asian, or Hispanic counterpart signals something about the trust relationship. This is a piece of information that subjects could read from the photograph and may be uncontrolled for by the attractiveness evaluations. We include dummy variables for the self-identified race or ethnicity of the photographs.[15]

Finally, we include several variables that may affect subject's assessments of attractiveness. For instance we control for whether the subject was smiling (almost 70 percent of the counterparts selected a smiling photograph), whether

[15] In the experiment subjects were asked to self-identify their gender and race or ethnicity. We used these categories to code subjects. It was clear from many of the evaluations that other subjects had a difficult time identifying the race or ethnicity of others. The same is true, in some instances, with identifying the gender of the subject. We have used rated ethnicity and gender as alternative forms of categorizing subjects, with similar results.

FIGURE 4
AVERAGE PERCENTAGE RETURNED BY WHETHER EXPECTATIONS WERE DASHED OR EXCEEDED



Averages are broken out by trusters (first-movers) who were one standard deviation above (attractive) or below (less attractive) the normalized mean rating of attractiveness. Standard errors are given by the vertical lines.

TABLE 4
DESCRIPTIVE STATISTICS FOR INDEPENDENT VARIABLES USED IN REGRESSION MODELS

	Trusters	Trustees
Mean-centered Attractiveness (mean, std)	.083 (.731)	.051 (.724)
Amount of Jewelry (mean, std)	–.111 (.608)	–.099 (.642)
Female 1st mover and Male 2nd mover (%, n)	25.23% (224)	27.25% (257)
Male 1st mover and Female 2nd mover (%, n)	22.75% (202)	23.44% (221)
Female 1st mover and Female 2nd mover (%, n)	21.85% (194)	18.56% (175)
Male 1st mover and Male 2nd mover (%, n)	30.18% (268)	30.75% (290)
African-American (%,n)	16.10% (143)	17.18% (162)
Asian (%,n)	18.24% (162)	18.66% (176)
Hispanic (%,n)	4.62% (41)	7.42% (70)
Smiling (%,n)	64.98% (579)	64.37% (607)
Glasses (%,n)	25.00% (222)	19.62% (185)
Dashed Expectations (%,n)	—	41.04% (387)
Total n	888	943

The data are broken out by trusters and trustees. Subjects were randomly assigned to images of counterparts who were not on-line. As a consequence the distributions vary for trusters and trustees. Trusters who sent nothing are excluded from the first column.

subjects were wearing glasses (which might obstruct a clear view of a person's eyes), and whether a subject was wearing a relatively large amount of jewelry. As to this latter item, evaluators were asked, on a scale of 1 to 6, how much jewelry was being worn. This included necklaces, earrings and piercings (a separate measure was also taken of the degree to which subjects had facial piercings, but because it is highly correlated with this measure it is excluded). The responses, like the attractiveness measure, were mean-centered for each individual evaluator.

The estimated model is a random effects regression and is presented in Table 5. The dependent variable is the amount of money sent by a truster. The attractiveness assessment is positive and statistically significant. This confirms, for the full data set, a beauty premium for trustees. More is sent to more attractive counterparts, even when controlling for a variety of characteristics of the counterpart. We do not find any differences in the amount sent for the sex pairings of the subjects. Likewise, we find no effect of a minority counterpart.[16] Smiling has a weak positive effect on contributions. Consistent with Scharlemann et al. (2001) we find that smiling invites trust. Whether the subject was wearing glasses has no effect. The amount of jewelry also has an effect, but it is weakly negative. This is surprising because if the amount of jewelry is a signal of status (or wealth) then the relationship is opposite of what we might expect. The conclusion we reach from this regression is that attractiveness matters for the trust decision independent of other characteristics of the counterpart.

What about on the reciprocity side? Here account for the expectations of the trustee as well as the factors used in the first-mover regression. Because subjects made multiple judgments we again use a random effects model for the analysis. The dependent variable is the percentage returned from the tripled amount sent by the truster. This allows us to make comparisons across reciprocators, no matter how much was sent. Those who were sent nothing were excluded from analysis as they were given nothing to return.

The primary independent variable is the attractiveness rating of the truster. If there is a beauty premium in the percent returned, then this coefficient should be positive. At the same time we control for male and female pairings with a series of dummy variables (with male/male pairings the omitted variable). Like the trust model we control for the race and/or ethnicity of the truster and a set of variables that might enhance or offset the attractiveness ratings.

Finally, we are concerned with the trustees' expectations. Table 3 suggests when expectations are not matched, the second-mover appears to punish the first-mover's lack of trust. Expectations were measured by asking subjects how much they expected the truster would send, prior to seeing what was sent. Dashed expectations are coded as a one when an expectation is not met and a zero otherwise. Figure 4 indicates that trusters who dash expectations are not punished uniformly, but rather it is the most attractive who are punished. To assess this an interaction term is calculated by taking the product of attractiveness and dashed expectations.

The second model in Table 5 provides estimates for the percentage returned. The first thing to note is that, while the coefficient on attractiveness variable is positive, its effect is insignificant. This implies that there is no beauty premium in reciprocity. However, it is clear that there are substantial effects due to dashed expectations. If a trustee's expectations are not met, then there is a substantial penalty assessed on the truster (amounting to more than 16 percent of what is returned on average). This penalty is larger for those who are more attractive. The interaction term between attractiveness and disappointment is strong and negative, connoting the

[16] In Eckel and Wilson (2003a) there are strong effects for race and ethnicity. The game used in that experiment has a different structure in which subjects make an all-or-nothing decision about trusting their counterpart. The sample of non-Caucasians in these data is too small to systematically study pairs of ethnic types in the same way in which the sex pairs were studied.

additional penalty for attractive first-movers. From this model it appears that the beauty penalty noted earlier results from an interaction between beauty, expectations, and willingness to punish a counterpart for dashed expectations.

There are some independent effects attributable to the control variables. In particular we find that female trustees return significantly more to male trusters. It is not apparent why this should be the case except to note that the signs of the coefficients for the female trustees are both positive and may reflect that females in general are more trustworthy. We find no effects for the race and ethnicity of the counterparts, although less is returned to Asians on average and this coefficient approaches significance. Smiles are positively related to returning more, but this effect is weak.

Generally, we find a beauty premium for attractive trustees. Those trustees appear to reciprocate that trust. In other words, attractiveness seems like a good bet for trust. On the other hand, this relationship is tempered by expectations on the part of the trustee. As we show, attractive trusters are penalized because they do not live up to the expectations held by their counterpart. Trustees expect attractive first movers to send more and they punish them when it doesn't happen. Our analysis of the extremes is supported in multivariate analysis of the full sample, controlling for other factors.

TABLE 5
RANDOM EFFECTS REGRESSIONS FOR THE AMOUNT SENT AND THE PERCENTAGE RETURNED IN THE TRUST GAME

	Model One Amount Sent	Model Two Percentage Returned
Constant	4.819***	33.089***
	(.404)	(2.158)
Mean-Centered Attractiveness of Counterpart	.178**	.839
	(.080)	(1.815)
Female 1st mover and Male 2nd mover	–.435	–.749
	(.578)	(2.235)
Male 1st mover and Female 2nd mover	.179	4.779**
	(.162)	(2.046)
Female 1st mover and Female 2nd mover	–.280	.462
	(.588)	(2.283)
African-American counterpart	.152	3.170
	(.149)	(2.269)
Asian counterpart	.084	–3.680*
	(.143)	(2.123)
Hispanic counterpart	–.319	2.944
	(.219)	(4.224)
Smiling counterpart	.224*	2.703
	(.135)	(1.736)
Counterpart wearing glasses	.079	.090
	(.140)	(2.288)
Amount of jewelry worn by counterpart	–.164*	–.201
	(.098)	(1.573)
Trustee's dashed expectations	—	–5.415***
		(1.675)
Interaction of truster attractiveness and dashed expectations	—	–4.886**
		(2.173)
	N = 943, r^2 = .03	N = 888, r^2 = .06

Note: *p < .10, **p < .05, ***p < .01
Standard errors are in parentheses. All subjects and subject pairings are included.

DISCUSSION AND CONCLUSION

These findings suggest we turn a deaf ear to our mothers' advice. Subjects not only trust strangers, but also they choose to trust based on a stranger's appearance. Several interesting findings arise from this research. First we find, in a controlled setting, that there is considerable trust and, on average, trust pays. This happens among strangers who are unlikely to ever interact in the future. It cannot be that the trust we see is a function of expected future interactions or an attempt to develop reputations. Nor does trust arise from a simple disposition on the part of the trusters. While there are some who always play the same strategy, over 75 percent of the first movers use a conditional strategy.

What is the basis for choosing a conditional strategy? Attractiveness appears to be an important factor. In line with bargaining experiments by Mulford et al. (1998), Solnick and Schweitzer (1999), Mobius and Rosenblat (2006) and Andreoni and Petrie (2003) we find that attractive people are treated differently. Attractive trustees are trusted more. In line with earlier work on stereotyping, we also find that beauty elicits positive expectations.

Beauty and expectations working in concert can undermine the beauty premium and produce a beauty penalty when positive expectations are dashed. Second movers have high expectations for more attractive counterparts. When those expectations are unfulfilled, attractive individuals are punished. Andreoni and Petrie (2003) find a similar pattern in their public goods game, and also note the relation to the literature on stereotyping. As additional information is made available about a counterpart (in their case the contributions made by others), this allows updating and diminishes or overturns the value of the stereotype. This is true for our second movers who form expectations and then have those expectations altered as they observe the behavior of the first mover.

Generally our findings are in line with the stereotyping literature. That literature shows that people use a stereotype to attribute characteristics to others. Our findings confirm what is found in political science: attractiveness matters for candidate evaluation [Ottati and Deiger (2002)]. At the same time our results extend to a different domain. We find that attributions about attractiveness extend to behavior in a laboratory setting with financial stakes. Indeed these attributions extend to basic trust relations. Others, like Hamermesh

and Biddle (1994), have noted that attractiveness often pays off in a financial way. Interestingly, we find that attractiveness does not always guarantee better treatment. We find a "beauty penalty" as well as a "beauty premium."

Why might these findings be important? First, for social scientists concerned with the role of trust in forming social capital it is useful to know that trust and trustworthiness are thriving among strangers. However, if trust builds social capital then those who are more attractive are more likely to be trusted in an initial interaction among strangers, and so have more opportunities to be successful in acquiring social capital. This suggests that attractive people may become more productive and so more valuable as exchange partners. Thus in the world outside the lab, there may be compelling reasons to trust the attractive. While it may seem silly to argue that attractiveness is the single or even the primary key to the accumulation of social capital, initial interactions (choosing whom to trust) have an important effect on subsequent interactions. Those who are not trusted at the outset have fewer opportunities to demonstrate that they can be trustworthy. This stops the process of building social capital (through establishing a reputation for trustworthiness) before it can even get started.

Second, there are important policy implications arising from evidence about who is and is not trusted. If people are willing to make potentially costly decisions on the basis of a brief assessment, then biases can develop in perceptions about the decisions of others. Potentially valuable interactions might be thwarted by the effect of biased expectations. Our work suggests that face-to-face interviews can extract important information from participants in an exchange. Inferences of trustworthiness may be more difficult in anonymous environment such as the internet, and more elaborate reputation-formation and contract-enforcement mechanisms may be necessary in such environments.

Third, these findings are important for reminding us that it is critical to account for systematic biases in strategic choice. While this setting might be conceived of as a signaling game [see for example Bianco (1998)] it would miss the fact that something deeper may be going on with constraints on human cognition. Paying closer attention to cognitive constraints and the ways in which perceptual biases enter into strategic behavior can help develop better models. A burgeoning sub-field in economics is taking both behavior and theory seriously. Coined "behavioral game theory" the aim is to use a subset of cognitive constraints to model strategic behavior. The primary statement comes from Camerer (1997) and Camerer (2003). Exemplars include Rabin (1993), Rabin (1998) and Loewenstein, O'Donoghue, and Rabin (2003). Much of this work considers the impact of psychology, including stereotyping and attribution as subsets. While early manifestations of behavioral economics tended to focus on anomalies or "silly" deviations from a narrowly defined concept of rationality, the later work has as its purpose a better understanding of human behavior, both within and outside the economic arena. Our findings point out that, in contrast to standard models of decisionmaking, a personal attribute such as attractiveness might be a useful signal of the potential quality of an exchange. A stereotype for attractiveness matches average behavior in the world. The human ability to make rapid judgments based on stereotypes means that decisions can be made quickly, economizing on the deliberate analysis of large quantities of information. No wonder we trust some strangers and make snap judgments about our books by their covers.

In closing, we return to the issue of the possible sources of a "beauty premium" or a "beauty penalty." While it is easy to say that a stereotype exists, it is much less clear to establish why attractiveness should constitute an important category for stereotyping. Attentiveness to attractiveness may be embedded as part of our cognitive apparatus and, if so, then it must be accounted for as an important part of our preferences. There are a number of evolutionary reasons for why humans might be attentive to attractiveness. It may signal good genes, substantial parental investment or status.

The good genes argument is straightforward. Attractiveness is comprised of a bundle of markers that highlight the genetic quality of an individual. These markers vary from hip-to-waist ratios to symmetry and averageness. Singh (1992) and Singh and Luis (1995) contend that a simple hip-to-waist ratio between .67 and .80 denotes attractiveness in females—largely because it is a marker of child bearing capacity, however, see the critique by Streeter and McBurney (2003). Others point to symmetry as a simple marker of gene expression. Humans (like most mammals) are notably bilaterally symmetric. Even minor deviations from symmetry are easily noticed and those deviations are commonly thought to represent "genes gone awry" in their expression. See Rhodes et al. (1998); Zebrowitz and Rhodes (2004)]. Average features are also found to be attractive. With respect to facial attractiveness this does not mean a commonplace visage, but rather a face that carries the average features for the population. Numerous studies have shown that "blended" faces (those that are compiled from many different images) are more attractive than single faces. See Rhodes et al. (2001); Valentine, Darling, and Donnelly (2004). Here too the argument is that deviations from the "average" may signal genetic mutation and as a consequence reflect bad genes. The good genes argument contends that both males and females have incentives to ensure high quality offspring. Under this conjecture both males and females would be attentive to the attractiveness of one another. The argument is largely one for mating purposes, but it seems reasonable that attention to attractiveness would slide over into other forms of behavior as well.

A second, but related, explanation focuses on parental investment. This argument assumes that the genetic material for most is of fairly high quality. Attractiveness remains an important marker, but is largely a function of clear skin, clear eyes, shiny hair, and symmetry. While good genetic material is a must and is linked to symmetry, many other outward features are related to diet and shelter. If resources are plentiful, then more is likely to be spent on children. However, when resources are scarce, parents often make decisions about which offspring are invested in and which

are not. While this is a seemingly cruel and heartless perspective, there is ample evidence that step-children are treated much differently than related children (Daly and Wilson 1988) and that humans discriminate amongst related children every bit as much as other animals when resources are scarce (Hrdy 1999, especially Chapters 19, 20). As a consequence, when individuals consider one another (again, largely for mating purposes) attractiveness is a marker of parental investment. It signals an estimate by parents as to which offspring will have the best chance of reproductive success. It may also signal which parents have the resources that can be invested in offspring—perhaps a doubly good signal of high quality genes.

The third argument derives from the first two and points to attractiveness as a marker for status. Here the claim is that humans, like most primates, live in dominance hierarchies. As such it is important to know who is near the top of the hierarchy and to be able to judge the relative status of individuals. It also makes sense to pay close attention to those who have greater status, if nothing else to emulate their behavior. After all, their position points to success (see for example Richerson and Boyd 2004). If attractiveness is partly genotypic and partly cultural, then those who seek higher status can over invest in culturally attractive trappings. If attractiveness is largely genotypic then status can be related to parental investment; parents with excess resources invest more heavily in their offspring. In either case, attractiveness is a marker that is correlated with status. As Webster and Driskell (1983) argue, attractiveness is a "diffuse status characteristic" in which those with it are treated differently and they, in turn, act differently because of how they are treated. In this regard the finding by Mueller and Mazur (1996) is striking. Their subjects evaluated and ranked photographs of West Point cadets, snipped from old yearbooks. The subsequent military career paths of those cadets were traced and their success was highly correlated with the evaluation of their photo. Status and attractiveness seem to go hand-in-hand, and humans appear to be very attentive to the latter. Unlike the good genes or parental investment arguments, the status argument does not necessarily point to attractiveness as purely a signal for mating.

The experiment reported here cannot sort between these three distal explanations of why attractiveness matters. This is a task for future research. It is clear that humans are attentive to attractiveness. Consequently attractiveness may not be a cheap talk signal, but instead may credibly signal some behaviors. These signals should be thought of as systematic aspects of human cognition and accounting for them might improve the predictive power of our models.

REFERENCES

Andreoni, James, and Ragan Petrie. 2004. "Public Goods Experiments without Confidentiality: A Glimpse into Fund-Raising." *Journal of Public Economics* 88 (7-8): 1605-23.

Bahry, Donna, and Rick K. Wilson. 2004. "Trust in Transitional Societies: Experimental Results from Russia." Paper presented at the American Political Science Asssociation, Chicago.

Berg, Joyce E., John W. Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10 (1): 122-42.

Berry, Diane S. 2000. "Attractiveness, Attraction, and Sexual Selection: Evolutionary Perspectives on the Form and Function of Physical Attractiveness." In M. P. Zanna, ed., *Advances in Experimental and Social Psychology*, pp. 273-342. San Diego, CA: Academic Press.

Bianco, William T. 1998. "Different Paths to the Same Result: Rational Choice, Political Psychology, and Impression Formation in Campaigns." *American Journal of Political Science* 42 (2): 1061-81.

_____. 1994. *Trust: Representatives and Constituents, Michigan Studies in Political Analysis*. Ann Arbor: University of Michigan Press.

Biddle, Jeff, and Daniel Hamermesh. 1998."Beauty, Productivity and Discrimination: Lawyers' Looks and Lucre." *Journal of Labor Economics* 16: 172-201.

Buchan, Nancy R., Rachel T. A. Croson, and Robyn M. Dawes. 2002. "Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange." *American Journal of Sociology* 108 (1): 168-206.

Burnham, Terence C. 2003. "Engineering Altruism: A Theoretical and Experimental Investigation of Anonymity and Gift Giving." *Journal of Economic Behavior & Organization* 50 (1): 133-44.

Burns, Justine. 2003. "Insider-Outsider Distinctions in South Africa: The Impact of Race on the Behaviour of High School Students." Paper presented at the Conference on Field Experiments in Economics, Middlebury College, Middlebury, VT, April 25-27.

Camerer, Colin F. 1997. "Progress in Behavioral Game Theory." *Journal of Economic Perspectives* 11 (4): 167-88.

_____. 2003. "Behavioral Game Theory: Experiments in Strategic Interaction." In Colin F. Camerer and Ernest Fehr, eds., *Roundtable Series in Behavioral Economics*. New York: Sage, and Princeton University Press.

Chaudhuri, Ananish, and Lata Gangadharan. 2002. "Gender Differences in Trust and Reciprocity." Paper presented at the American Economic Association Meetings. Atlanta, GA.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94 Supplement S95-S120.

Croson, Rachel, and N. Buchan. 1999. "Gender and Culture: International Experimental Evidence from Trust Games." *American Economic Review* 89 (2): 386-91.

Daly, Martin, and Margo Wilson. 1988. *"Homicide." For Foundations of Human Behavior*. New York: de Gruyter.

Eagly, A. H., R. D. Ashmore, M. G. Makhijani, and L. C. Longo. 1991. "What Is Beautiful Is Good, But: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype." *Psychological Bulletin* 110:109-28.

Eckel, Catherine C., and Rick K. Wilson. 2003a. Conditional Trust: Sex, Race, and Facial Expressions in a Trust Game. Paper presented at the Trust and Institutions, Harvard University, April 24-26.

_____. 2003b. "The Human Face of Game Theory." In Elinor Ostrom and James M. Walker, eds., *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, pp. 245-74. New York: Sage.

_____. 2004. "Is Trust a Risky Decision?" *Journal of Economic Behavior & Organization* 55 (4): 447-66.

_____. 2006. "Internet Cautions: Experimental Games with Internet Partners." *Experimental Economics* 9 (1): 53-66.

Ekman, Paul. 2003. "Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life." *Times Books*. New York: Henry Holt.

Feingold, Alan. 1992. "Good-Looking People Are Not What We Think." *Psychological Bulletin* 111 (2): 304-41.

Fenno, Richard F., Jr. 1978. *Home Style: House Members in Their Districts*. Boston, MA: Little, Brown.

Fershtman, Chaim, and Uri A. Gneezy. 2001. "Discrimination in a Segmented Society: An Experimental Approach." *Quarterly Journal of Economics* 116 (1): 351-77.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91 (3): 481-510.

Hamermesh, Daniel, and Jeff Biddle. 1994. "Beauty and the Labor Market." *American Economic Review* 84: 1174-94.

Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Sage.

Hayashi, N., E. Ostrom, J. Walker, and T. Yamagishi. 1999. "Reciprocity, Trust, and the Sense of Control—a Cross-Societal Study." *Rationality and Society* 11 (1): 27-46.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92 (5): 1644-55.

Hrdy, Sarah Blaffer. 1999. *Mother Nature: A History of Mothers, Infants, and Natural Selection*. New York: Pantheon.

Jackson, Linda A., John E. Hunter, and Carole N. Hodge. 1995. "Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review." *Social Psychology Quarterly* 58: 108-22.

Karlan, Dean. 2003. Using Experimental Economics to Measure Social Capital and Predict Financial Decisions. Paper presented at the Conference on Field Experiments in Economics, Middlebury College, April 25-27.

Koford, Kenneth.2001. Cooperation in Horizontal and Vertical Relationships: Bulgaria and the West. Paper presented at the 20th Arne Ryde Symposium on Experimental Economics, Lund, Sweden, November 9-11.

Langlois, Judith H., Lisa Klakanis, Adam J. Rubenstein, Andrea Larson, Monica Hallam, and Monica Smoot. 2000. "Maxims or Myths of Beauty? A Meta-Analysis and Theoretical Review." *Psychological Bulletin* 126 (3): 390-423.

Levi, Margaret, and Laura Stoker. 2000. "Political Trust and Trustworthiness." *Annual Review of Political Science* 3 (1): 475-507.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4): 1209-48.

Mobius, Markus M., and Tanya S. Rosenblat. 2006. "Why Beauty Matters." *American Economic Review* 96: 222-35.

Mueller, Ulrich, and Allan Mazur. 1996. "Facial Dominance of West Point Cadets as a Predictor of Later Military Rank." *Social Forces* 74: 823-50.

Mulford, Matthew, John Orbell, Catherine Shatto, and Jean Stockard. 1998. "Physical Attractiveness, Opportunity, and Success in Everyday Exchange." *American Journal of Sociology* 103 (6): 1565-92.

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

_____. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

Ostrom, Elinor, and James M. Walker, eds. 2003. *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*. New York: Sage.

Ottati, Victor C., and Megan Deiger. 2002. "Visual Cues and the Candidate Evaluation Process." In Victor C. Ottati, R. Scott Tindale, eds., *Social Psychology of Politics*, pp. 75-87. New York: Kluwer Academic/Plenum.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83 (5): 1281-302.

_____. 1998. "Psychology and Economics." Journal of Economic Literature 36 (1): 11-46.

Rhodes, Gillian, Fiona Proffitt, Jonathon M. Grady, and Alex Sumich. 1998. "Facial Symmetry and the Perception of Beauty." *Psychonomic Bulletin & Review* 5 (4): 659-69.

Rhodes, Gillian, Sakiko Yoshikawa, Alison Clark, Kieran Lee, Ryan McKay, and Shigeru Akamatsu. 2001. "Attractiveness of Facial Averageness and Symmetry in Non-Western Cultures: In Search of Biologically Based Standards of Beauty." *Perception* 30 (5): 611-25.

Richerson, Peter J., and Robert Boyd. 2004. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Scharlemann, Jorn P. W., Catherine C. Eckel, Alex Kacelnik, and Rick K. Wilson.2001. "Value of a Smile: Game Theory with a Human Face." *Journal of Economic Psychology* 22 (5): 617-40.

Singh, Devendra. 1992. "Body Shape and Women's Attractiveness: The Critical Role of Waist-to-Hip Ratio." *Human Nature* 4 (3): 297-321.

Singh, Devendra, and Suwardi Luis. 1995. "Ethnic and Gender Consensus for the Effect of Waist-to-Hip Ratio on Judgment of Women's Attractiveness." *Human Nature* 6 (1): 51-65.

Solnick, Sara J., and Maurice E. Schweitzer. 1992. "Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions." *Organizational Behavior and Human Decision Processes* 79 (3): 199-215.

Streeter, Sybil A., and Donald H. McBurney. 2003. "Waist-Hip Ratio and Attractiveness: New Evidence and a Critique of a 'Critical Test'." *Evolution & Human Behavior* 24 (2): 88-98.

Uslaner, Eric M. 2002. *The Moral Foundations of Trust*. Cambridge, MA: Cambridge University Press.

Valentine, Tim, Stephen Darling, and Mary Donnelly. 2004. "Why Are Average Faces Attractive? The Effect of View and Averageness on the Attractiveness of Female Faces." *Psychonomic Bulletin & Review* 11 (3): 482-87.

Webster, Murray, and James E. Driskell. 1983. "Beauty as Status." *American Journal of Sociology* 89 (1): 140-65.

Yamagishi, Toshio, Karen Cook, and Motoki Watabe. 1998. "Uncertainty, Trust, and Commitment Formation in the United States and Japan." *American Journal of Sociology* 104 (1): 165-94.

Zebrowitz, Leslie A., and Gillian Rhodes. 2004. "Sensitivity to 'Bad Genes' and the Anomalous Face Overgeneralization Effect: Cue Validity, Cue Utilization, and Accuracy in Judging Intelligence and Health." *Journal of Nonverbal Behavior* 28 (3): 167-85.

Zuckerman, Marvin. 1994. *Behavioral Expression and Biosocial Bases of Sensation Seeking*. New York: Cambridge University Press.

Received: August 5, 2004
Accepted for Publication: August 9, 2005
rkw@rice.edu
eckelc@utdallas.edu