Motivation and Emotion, Vol. 18, No. 2, 1994

Trust, Social Categories, and Individuals: The Case of Gender[1]

John Orbell
University of Oregon

Robyn Dawes
Carnegie-Mellon University

Peregrine Schwartz-Shea
University of Utah

While the cooperate vs. defect choice in the prisoner's dilemma is not an appropriate paradigm for the study of trust and trusting behavior, the play vs. not play choice is. We show that females as a category *are more trusted to cooperate — by both male and female judges — than males. Yet neither male nor female judges use gender to predict cooperation from particular individuals (trust) or as a criterion for choosing to play (trusting behavior) when they have the option of not playing particular prisoner's dilemma games. Further, in our experimental context, female and male did not differ in their cooperation rates. We speculate (a) that subjects' generalized expectations are a response to gender-based role differences outside the laboratory, (b) that subjects' failure to make individual-by-individual discriminations by gender is a response to the fact that the experimental context made such natural-world roles irrelevant, and (c) that our findings about the irrelevance of gender* per se *in trusting relationships will be true for other social categories* per se.

[1] The research reported in this paper was supported by the National Science Foundation grant SES-9008157 to Orbell and Dawes; any opinions, conclusions or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation. Pamela Ferrara played an important role in administering all stages of the experiments; Jean Stockard gave us a helpful methodological and substantive critique; and Debbie Elvin, Tom Morikawa, Matthew Mulford, and Catherine Shatto each helped at several stages of the paper.

Trust, as we all know, can be *earned.* If, through a long relationship, you have never exploited me — despite many opportunities for doing so — then I can be confident that you will not take advantage of me *next* time you have the opportunity to do so. Sometimes, however, we have to decide whether to trust strangers, individuals with whom we have no experience, or even knowledge. Then *categories* may be useful. To what category does this stranger belong? What is to be expected from members of that category *in general?* The categories having a reputation for trustworthiness can provide a basis for my trusting this *particular* stranger.

Our interest is in the trust that people accord members of the categories "male" and "female," and in the extent to which that trust guides their decisions in practice. Do people trust one gender more than the other, and does their *categoric* trust lead them to act in a more trusting manner with particular males and particular females?

Defining Trust and Trusting Behavior

Minimally, *trust* involves the expectation of some desirable behavior from another person. I trust you to do *a* rather than *b* if I assign a probability greater than .5 to your doing *a;* and I trust you completely if I assign a probability of 1.0 to that.

There is more to trust than expectations, of course. We expect the sun to rise, but we do not trust it to do so; in everyday use, the trusted person (the target) is a choosing agent. Further, in everyday use the trusting person (the judge) *prefers* the target to do *a,* and the target has some *incentive* or reason to do *b.* We do not say "I trust you to do (something I don't want)" and neither do we say "I trust you to do (something you would have done anyway)."

Also, trust implies vulnerability. The judge has a choice to make, and payoffs from that choice are contingent on the target's choice. It follows that we can distinguish *trust* from *trusting behavior.* The former involves expectations in the terms discussed above, and the latter involves the judge's action on those expectations. One might have a relatively high level of trust, but — because of the costs of being wrong — one might, nevertheless, choose to act in an untrusting way. By hypothesis, then, trusting behavior is the product of an expected value calculation involving both expectations and the values at stake.

The prisoner's dilemma has some features that make it an attractive paradigm for studying trust. Both parties in that game must make a choice, and both are vulnerable in their choices; each player's payoff is dependent, not only on his or her own action but on the other individual's action as well. Each player confronts a temptation that provides a reason for the other's doubt that he or she will cooperate. Finally, the fact that mutual cooperation provides a "cooperative surplus" (Gauthier, 1986) is consistent with the intuition that trusting behavior is socially productive.

Yet the classic dilemma is not in itself a satisfactory paradigm for the study of trusting behavior. As Riker (1980) put it:

> Unfortunately in the prisoners' dilemma the alternative of trusting is strictly dominated by the alternative of self-reliance, so trust is rendered immediately and logically irrational, unless the dilemma is resolved. Thereby trust is also rendered unpredictable. (p. 11)

In other words, an appropriate framework for studying trust would make both trusting behavior and nontrusting behavior rational — under some circumstances. But the fact that cooperation is, by definition, dominated in the prisoner's dilemma, means that trust-as-cooperation can *never* be rational.[2]

The choice between playing and not playing given prisoner's dilemma games can, however, be usefully addressed in rational choice terms. As several authors have pointed out recently, in natural situations, we often do have the choice between playing and not playing (Orbell & Dawes, 1991, 1993; Orbell, Schwartz-Shea, & Simmons, 1984; Schuessler, 1989; Tullock, 1985; Vanberg & Congleton, 1992; Yamagishi, 1988). Clearly, trust is intimately involved in our decisions between playing and not playing prisoner's dilemma games with other people. And the rationality of making that choice one way or the other can depend critically on expectations about the other's behavior.

Consider the structure

$$t > c > 0 > d > s \qquad (1)$$

where, conventionally, t is the free rider's payoff, c is the payoff for mutual cooperation, d is the payoff for mutual defection, s is the sucker's payoff, and zero is the payoff for not playing the game. In this case, whatever one's own intentions between cooperation and defection, it is rational to play if a potential partner is expected to cooperate, and not to play if a potential partner is expected to defect.

Note that, as in all such calculations, the magnitudes at stake can also be critical. Consider, for example, the prisoner's dilemma matrix

[2] Prisoner's dilemmas that are iterated may be conceptualized as involving trust, but by their very nature (Rapoport, 1967) they are not classical prisoner's dilemmas, because players may influence other players' choices across the iterated trials.

$$\begin{array}{cc} 2,2 & -4,6 \\ 6,-4 & -1,-1 \end{array}$$

with 0 as the payoff for not playing. If I intend cooperation and assign a probability of .6 to your also cooperating — thus a .4 probability to your defecting — my expected value from playing is –.40. That is less than 0, thus, rationally, I should not play. Alternatively, if I intend defection, my expectation from playing is 3.2. Thus, rationally, I should play. Defining "trust" as expectations of a potential partner's cooperation, and "trusting behavior" as willingness to play the game, are consistent with the everyday understanding that neither entering nor refusing to enter possibly profitable but risky relationships is necessarily irrational.[3]

Trust-as-expectations will not always be decisive for rational play versus not-play choices. For example, with

$$t > c > d > s > 0 \tag{2}$$

one should always play regardless of one's trust in the other's cooperation (playing is dominant), and with the structure

$$0 > t > c > d > s \tag{3}$$

one should never play regardless of such trust (playing is dominated).

Equally, some of the everyday meanings of trust and trusting behavior are, no doubt, not captured by this framework. But a surprising number are,[4] and a great advantage of this formulation is that it permits systematic,

[3] Of course, entering and choosing to cooperate remains dominated by entering and choosing to defect. Our analysis of trust does not explain a choice to cooperate, once the choice to play has been made. That choice requires something other than a rational response to prisoner's dilemma incentives, for example, iteration of the game, evolved social norms, ethical constraints, or any of the other mechanisms discussed in the extensive literature. We are discussing rationality in the context of the choice between entering versus not entering, not in the context of the trinary choice between entering and cooperating, entering and defecting, and not entering. Absent such "supplementary incentives," entering and defecting, of course, dominates entering and cooperating.

[4] In his critique of those who study trust via the cooperate versus defect choice alone, Riker (1980) cited (a) buyers and sellers of a house agreeing on a price and trusting each other to keep that agreement; (b) one member of a marriage who is confronted by a spouse's interest in a third party but who has, nevertheless, a dominant incentive to trust the spouse's honorable behavior; (c) von Neumann and Morgenstern's paradigm n-person game in which three players must choose one of the other two in the hope that the chosen one reciprocates. We agree that these examples all transcend the prisoner's dilemma, and that "trust," in its everyday meaning, is involved in all three. Yet each does involve the dependency, possibility of gain, temptation to breach the other's trust, and — most important — freedom to enter

laboratory testing of hypotheses about trust and trusting behavior. We present some such hypothesis testing in this paper.

A contractarian might argue that trust should be understood as the expectation that a person will fulfill his or her side of a more-or-less explicit bargain, promise, or contract. But, as Baier (1985) pointed out, contract implies the threat of coercion for noncompliance, thus "very minimal trust."

> [Contract] is an interesting case of the allocation of trust of various sorts, but it surely distorts our moral vision to suppose that *all* obligations, let alone all morally pressured expectations we impose on others, conform to that abnormally coercive model.

We prefer trust as expectations about cooperation because it is a broader conceptualization, one that includes — but is not limited to — prisoner's dilemmas in which there has been some prior agreement. Trust can be a critical issue with respect to strangers, no less than with respect to those with whom we have a prior agreement about how we should behave.

Nevertheless, the possibility of contracts, promises, and agreements does raise the general issue of the criteria by which we come to trust others — and, thus, are willing to enter prisoner's dilemma games with them. Contracts, promises, and agreements do, usually, increase our expectation that others will cooperate, no doubt all the more so when they are backed up by a credible legal system. But each requires at least some interaction prior to decision making (Orbell, Dawes, & van de Kragt, 1989), and that is not always possible. Absent such a history, on what do we base our trust?

Elsewhere (Orbell & Dawes, 1991; 1993), we have proposed, as one criterion, that people project to others' intentions from their own intentions. Riker (1980) calls this *introspection* and proposes, in addition, that people use (a) learning — the experience they gather from having interacted with individuals (and, presumably, populations) previously; (b) utilitarianism — recognition, for example, that potential partners confront dominant strategies; and (c) rules of thumb — the adage, for example, to "never trust anyone over 30."

We are interested in the importance of categorical thinking in trust and trusting behavior. What is the importance of *social categories* in determining the trust that judges assign to individuals? In particular: *Does the gender of a potential partner serve as a basis for trust and for trusting behavior?*

or not enter a relationship that our definition of trust requires. Remembering that trust need *not* be decisive in entering prisoner's dilemmas (viz., the location of the "not play" might make trust irrelevant), each of these examples might, quite satisfactorily, be reformulated in our terms.

GENDER, TRUST, AND TRUSTING BEHAVIOR

There have been no empirical studies of gender as a basis for choosing partners in prisoner's dilemma games, but there are studies of the relationship between gender and cooperation: Their findings are inconclusive. Some laboratory studies (e.g., Aranoff & Tedeschi, 1968; Eagly & Crowly, 1986; Jones, Steele, Gahagan, & Tedeschi, 1968; Meux, 1973) report more cooperation among women, whereas others (e.g., Brown-Kruse & Hummels, in press; Kahn, Hottes, & Davis, 1971; Rappoport & Chammah, 1965) report more among men. Still others (e.g., Dawes, McTavish, & Shaklee, 1977; Javine, 1986; Stockard, van de Kragt, & Dodge, 1988) report no difference — or only slight and inconsistent differences.

Gilligan's well-known (1982) argument that women are more "contextual" in their moral thinking, and that men are more "instrumental," has been interpreted by some as predicting more cooperation from women than from men in prisoner's dilemma games (e.g., Brown-Kruse & Hummels, 1993; Sell, Griffith, & Wilson, 1991). This has intuitive appeal, but her case has not been supported well by evidence from investigations of ongoing behavior (Hamilton, 1986; Stockard & Johnson, 1992). Further, Gilligan's methodology has been criticized: Her conclusions are based on self-reported, retrospective evaluations that are highly susceptible to social stereotyping (Pearson, Ross, & Dawes, 1992). And, as Brown-Kruse and Hummels (1993, p. 13) have pointed out, people often "fail to put their money where their mouth is," talking cooperatively but not behaving in a cooperative manner when actual choices must be made. The point is, of course, reminiscent of La Pierre's (1934) classic finding that peoples' generalized dispositions with respect to groups predict only poorly their actual choices with respect to individuals—a point to which we return in our discussion section.

Finally, the informality of Gilligan's argument makes the translation to prisoner's dilemma terms problematic. "Contextual" and "instrumental" are motivational categories for Gilligan, not behavioral categories. It requires only modest theoretical ingenuity to invent hypotheses specifying contextual reasons that might lead people to defect, or instrumental reasons that might lead them to cooperate.

To anticipate, within our experimental paradigm we find that women are expected — are trusted — to cooperate more than men, and that this expectation is held equally for male and female judges. But this is so only for generalized expectations; gender is not a basis for predicting cooperation from particular males and particular females in particular situations.

We find, further, that *trusting behavior* — choosing actually to enter particular prisoner's dilemma relationships — is not gender-based. At least in the laboratory, people do not base decisions between entering and not entering prisoner's dilemma games on the gender of potential partners.

In our paradigm, *trustworthiness is* equated with cooperativeness per se, and our laboratory data support those earlier findings that show no gender differences in cooperativeness — meaning that our subjects were incorrect in their generalized expectations, but were correct in their failure to discriminate among potential partners by gender. In a concluding section we speculate that the myth of greater female cooperativeness has arisen in response to the gender-biasing of social roles beyond the laboratory.

EXPERIMENTAL DESIGN

Six subjects, recruited by advertisements in a student newspaper and a local daily,[5] were seated face-to-face around the periphery of a large room. The basic structure required each subject to make one decision interacting with each of the other five who were present. Subjects' five decisions involved the five matrices listed in the Appendix. These matrices were designed to meet the requirements: (a) $t > c > 0 > d > s$; (b) that a subject who captured the free rider's payoff on all five occasions would gain $20; and (c) that a subject who was suckered on all five occasions would lose $20. Note that Matrices *c* and *e* are the same.

In Experiment 1 subjects made the standard binary choice between cooperating and defecting; they had no option but to play with each of the other five subjects. In Experiment 2, however, subjects had, in addition, the alternative of "opting out." If either subject chose this alternative, both would make nothing from that particular interaction — no gain, but also no loss. Subjects' take-home pay was the sum of their payoffs across all of their interactions with the five others present.

To have any bite, the opt-out alternative required that subjects could lose money, but we could not, of course, take their own money from them. Our solution was to pay them $20 for participation in an initial study with the understanding that, in the subsequent decision-making study, they could lose all that money — or double it. The initial study required subjects to read a number of politicians' statements from the Oregon Voter's Pamphlet and to respond to them through an extended questionnaire. At the con-

[5] Initial contact was made by telephone, at which time they were assigned to time slots by their convenience (with every effort being made to schedule couples or people who knew each other at different times); time slots were subsequently assigned to experimental conditions and replications randomly with no effort to organize by gender (or any other criterion). The only constraints on subjects' signing up were age (we did not accept anyone under 18 years old) and having participated in the experiment before.

clusion of this earlier study (which took about 40 minutes), they were handed the $20 in singles, they counted it, put it in plastic bags, and they carried those bags to a further room down a corridor where the subsequent study would be conducted. The bags each had a letter between A and F on them, designations that determined subjects' seating and identification (ID) number in the subsequent experiment.

Subjects placed their money bags on a table in the center of this subsequent room, the money remaining in full view for the duration of the second experiment. Instructions made it clear that this $20 would be their "starting money" for the second study, and that ("depending on your decisions and the decisions of others here") they could as much as double it or lose it all — or end up with something between those extremes.

Undoubtedly subjects' fear of losing money earned in this way would not be as compelling as their fear of losing money that they had brought to the laboratory with them. We believe, however, that this device did make subjects take the prospect of loss more seriously than had we just handed them $20 — or asked them to imagine the possibility of loss.

At several points, the instructions also assured subjects that their choices would be completely anonymous. Choices would be known to the experimenters who read them into the computer at the conclusion of the experiment, but this knowledge was a constant factor across all conditions and choices. Decisions would be recorded using clipboards so that others could not see what they were writing. Subjects would leave the experiment room for a pay room one-by-one, and each would be well clear of the general area before the next was released. They were also told (truthfully) that there was no deception in the study, that it was important they understood everything that was going on, and that they should ask questions whenever anything in the instructions was not clear.

After the instructions were read, we gave subjects a brief quiz testing their understanding. Answers were reviewed and any part of the instructions that caused problems was explained again. The experimenter did not proceed until satisfied that everyone did, in fact, understand completely. (Notice that there was no discussion among subjects about the choices that they faced with respect to each other.)

Decisions were recorded on five "decision forms" which were turned over one-by-one and in unison. Each of these forms had on it (a) the ID of one of the other subjects; (b) the payoff matrix for the particular interaction; (c) a place for recording the decision in the particular case — either X (cooperate), Y (defect), or additionally in the trinary experiment, O (for opt out); and (d) an 11-point scale for recording expectations about others' X versus Y choices and, in the case of the trinary experiment, their play versus opt-out choices.

Instructions about how to indicate expectations on the 11-point scale in the binary experiment were as follows:

> Now look at the place for recording your expectations about what the other person is going to do. You will put a mark in one of the eleven boxes; in general, the closer to the right, the more confident you are that the other person will choose Y [viz., will cooperate]; the closer to the left, the more confident you are that the other person will choose X; and if you think it is a tossup, you will put a mark in the middle (50/50) box.

On the decision form, the wording was "Your best estimate about what person _____ will choose is: _____" and the scales themselves were labeled with "Certain X" at the extreme left, and "Certain Y" at the extreme right.

In the trinary experiment with the option of not playing, subjects were instructed to estimate expected cooperation from the other person in the event he or she were to play. They were also instructed to record their expectations of the other person's opting out versus playing in an similar manner. (In the trinary experiment we made it plain that subjects should record their expectations of another's cooperating versus defecting even if they were certain that he or she would not, in fact, choose to play.)

The Appendix shows the five matrices used, and the sequence in which the subjects played with each other on those matrices. Note that, while given pairs of subjects did not play with each other at the same time, play between them always involved the same matrix.

It would have been logistically impossible for all pairs to play with each other on the same move. As it turned out, it was also fortuitous that they did not; subjects characteristically "checked out" their prospective partner in each case quite carefully, and both members of a pair doing that at the same time would have permitted passage of implicit or explicit cues and assurances between them.

Once the decision forms were completed, an experimenter took them to the payout room where decisions were read into a payout program. While that was happening, subjects filled in one questionnaire asking for personal information, and another asking for estimates about how individuals in various categories (including male and female) would make choices such as the ones they had just made. This second questionnaire referred subjects to the prisoner's dilemma matrix in Table I and instructed them as follows:

> Here we want your best possible estimate about what decisions different people — people in general, not just those in this room — would make if they were choosing with the payoffs specified on the cover sheet. When you make these estimates, assume that the people are making their decisions in *exactly* the same circumstances you just did.

Table I. The Prisoner's Dilemma Matrix to Which Subjects were Referred When Recording Their Generalized Expectations

	If the other chooses	
If you choose	X	Y
X	YOU gain $4 OTHER gains $4	YOU lose $8 OTHER gains $8
Y	YOU gain $8 OTHER loses $8	YOU lose $4 OTHER loses $4

As with their individual-by-individual expectations, these were recorded using a 10-point scale from "certain X" to "certain Y" and (in the trinary experiment) from "certain stay" to "certain opt out."

We ran 18 six-person groups (108 subjects) in both the binary and the trinary experiments.

FINDINGS

Gender and Expected Cooperation

Table II reports, by gender of judge and for each experiment, mean expectations about males and females "in general," and about particular males and particular females in the judge's experimental session. For analytic purposes, in the latter case we have averaged judges' expectations for all a subject's partners who were male and for all who were female. Then, for each experiment, we have conducted a repeated measures analysis of variance where the factors are expectations about males and females in general, expectations about the other males and females in the group—particular individuals—and the between-subjects variable is judges' gender.

The results are the same in both experiments. First, particular individuals are expected to be more cooperative than people in gender categories: for the binary experiment, $F(1, 90) = 13.38, p < .001$; for the trinary experiment, $F(1, 99) = 20.74, p < .001$. Note that this comparison is not perfect: The payoffs in the matrix to which subjects were responding in their generalized expectations were different from those in the matrices to which they were responding in their case-by-case decision making.

Table II. Mean Expectations About Cooperation for Males and Females "in General," and for Particular Males and Particular Females

Judge	Males in general	Females in general	Particular males	Particular females
		Binary experiment		
Male	.396	.528	.522	.593
Female	.407	.593	.558	.589
		Trinary experiment		
Male	.420	.590	.629	.576
Female	.315	.615	.506	.576

Second, females are expected to cooperate substantially more than males: for the binary experiment, $F(1, 90) = 34.00, p < .001$; for the trinary experiment, $F(1, 99) = 34.89, p < .001$. Third, the expectation about greater female cooperativeness is much stronger for generalized gender categories (men and women in general) than for particular individuals: for the binary experiment, $F(1, 90) = 8.26, p < .005$; for the trinary experiment, $F(1, 99) = 35.69, p < .001$.

In addition, the gender of the judge had just one significant effect in either experiment. In the trinary experiment only, women had a greater discrepancy for female versus male targets than did men, $F(1, 99) = 9.40$, $p < .005$. This finding was *not* replicated, however, in the binary experiment, $F(1, 90) = 0.03$.

The interaction between judgments about gender categories and judgments about particular others does not, in itself, imply that judgments about particular targets do not differ as a function of target gender. To test whether they do, we conducted post hoc ANOVAs that simply omitted judgments about the categories men and women "in general." The results were inconclusive. In the Binary experiment particular women were expected to be more cooperative than were particular men, $F(1, 93) = 8.204$, $p = < .005$, but in the Trinary experiment gender of target made no difference, $F(1, 99) = 0.124, p > .5$. Gender of judge made no difference in either case; nor were any of the interactions significant.

We conclude that people — male and female alike — do expect females to cooperate in prisoner's dilemma games more than males, but that this expectation is more strongly based on gender categories than on gender differences between individual targets. There is a strong categoric expectation that does not carry over to expectations about particular individuals in particular circumstances.

Gender and Expected Opting-Out

In studying expectations about others' willingness to play these games we are, of course, limited to data from the trinary experiment. In Table III we report, by gender of judge, subjects' mean expectations about the play versus no-play choices of "males in general" and "females in general," and averaged expectations across all partners who were male and all partners who were female.

Once again, we conducted a repeated measures analysis of variance. In this case, the factors are play versus not-play expectations about males and females in general and about other males and females in the group, and the between subjects variable is judges' gender. We find, first, that particular people are expected to play substantially more than are people in general, $F(1, 73) = 32.252, p < .001$; second, that males are expected to play more than females, $F(1, 73) = 52.851, p < .001$; third, that there is no effect by gender of judge, $F(1, 73) = 0.117$.

No interaction is significant, but that fact does not preclude effects within the two judgment categories. To test that possibility, we conducted two separate post hoc ANOVAs. There is no effect by gender of judge in either case, "in general" $F(1, 106) = 1.976, p = .163$; "in particular" $F(1, 73) = 0.292$, but the interaction was significant in the "in general" case, $F(1, 106) = 4.050, p = .047$. Women expected men "in general" to play relatively more than men expected them to play.

We conclude that expectations about entering these games are unambiguously based on the gender of the target. Distinct from expectations about cooperation — which we equate with trust — this finding extends *both* to generalized expectations and to expectations about particular individuals.

Trusting Behavior

The initial issue is peoples' willingness actually to enter these relationships — behavior we recognize as trusting. To examine that, we turn

Table III. Mean Expectations About Staying and Playing for Males and Females "in General," and for Particular Males and Particular Females: Trinary Experiment

Judge	Males in general	Females in general	Particular males	Particular females
Male	.655	.467	.936	.655
Female	.743	.486	.908	.622

to subjects' five play versus not-play choices in the trinary experiment. For analytic purposes, we computed the proportion of each subjects' encounters with females and with males in which the decision to "play" was made.[6] Table IV reports the mean proportion of "play" choices by gender of judge and gender of target.

We conducted a repeated measures analysis of variance with proportion of "play" choices with male and with female as the within-subjects factor and the subject's gender as the between-subjects factor. Despite their generalized expectation that females would cooperate more than males, we find no difference in our subjects' willingness actually to play with females and males, $F(1, 99) = 2.513, p = .116$, and no difference in this respect by gender of judge, $F(1, 99) = 0.371, p = .544$. Neither was there an interaction between gender of judge and willingness to play with male or female, $F(1, 99) = 0.024$.

We conclude that, in the laboratory at least, our subjects — male and female alike — do not base their willingness to trust on targets' gender. Despite their generalized expectation of female cooperativeness, their actions *in practice* are gender blind.

Trustworthy Behavior

What of possible gender differences in actual cooperation? To explore this much-discussed issue, we constructed for each subject the proportion of cooperative choices among his or her encounters with females, and the proportion of cooperative choices among his or her encounters with males. Table V reports, for the two experiments, the proportion of cooperative choices by gender of judge and gender of target.

Table IV. Proportion of "play" Choices by Gender of Judge and Gender of Target: Trinary Experiment

	Proportion of "play" choices		
Judge	With males	With females	Diff.
Male	.658	.704	.046
Female	.596	.686	.070

[6] Of course, the number of encounters with males and with females could vary between 1 and 5; when there were no others of a particular gender in a group, the case was recorded as missing data.

Table V. Proportion of "Cooperate" Choices by Gender of Judge and Gender of Target

	Proportion of "cooperate" choices	
Subject	With males	With females
	Binary experiment	
Male	.417	.415
Female	.448	.519
	Trinary experiment	
	Proportion of (play and) "cooperate" choices	
	With males	With females
Male	.557	.520
Female	.517	.587

(If there were no others of a given gender in the group, the case was recorded as missing data. In the binary experiment, of course, all five encounters required subjects either to cooperate or to defect, but in the trinary experiment subjects could also refuse to play — as many did. For the trinary experiment, therefore, the base from which we constructed these variables is the proportion of cooperate choices in those instances in which the subject *did not* opt out. Those instances in which the subject *did* opt out were recorded as missing data.)

We find no difference in cooperativeness by subject's gender, for the binary experiment: $F(1, 93) = 0.844$, for the trinary experiment: $F(1, 68) = 0.191$; no difference by gender of target, for the binary experiment: $F(1, 93) = 0.819$, for the trinary experiment: $F(1, 68) = 0.039$; and no interaction in this respect between subject's gender and gender of target, for the binary experiment: $F(1, 99) = 0.895$, for the trinary experiment: $F(1, 68) = 0.250$.

We conclude that our subjects are broadly correct in their failure to discriminate by target gender in their individual-by-individual expectations: Members of both genders cooperate at statistically indistinguishable rates whatever the gender of their partners. Further, the fact that the pattern shows up in the trinary data lets us conclude that the "screening" effect of the opt-out alternative makes no difference in this respect.

Payoff Differences by Gender?

Rose (1992) has argued that (a) if women have a greater "taste for cooperation" than men; and (b) if women "are merely *perceived* to have a greater taste for cooperation than men," they will be disadvantaged relative to men. If the first assumption holds, men can — for example — offer less favorable terms to women when negotiating prisoner's dilemma relationships, and womens' cooperativeness will lead them to accept. If the second holds, employers — for example — will make lower offers to women than to men, finding support for their expectation in the fact that some women, at least, will accept such low offers.

A focus on the option of not playing such games might lead to the opposite conclusion. If women were generally believed to be more cooperative than men, women would make and accept offers of play disproportionately with each other; and if they were more likely to cooperate than men, they would benefit from such relationships, as well as from the systematic avoidance of less cooperative males. Men, on the other hand, would be relatively excluded from the fruits of interacting with the relatively more cooperative female population.

Because male and female cooperate at about the same rates, however, even if peoples' play versus no-play choices were based on a myth of female cooperativeness (which they are not), neither gender would be relatively advantaged. Even if there were more consummated plays involving women, the proportion of cooperative encounters would be unaffected.

And, in fact, mean payout for male and female in our experiments was almost identical. In both our experiments, and across all five decisions, the average payout to males was $20.60, and the average payout to females was $20.70. Prisoner's dilemma play returned subjects a (very) modest profit above the $20 with which they entered the experiment, but it returned that equally to male and female.

DISCUSSION AND CONCLUSIONS

The failure of generalized gender-based expectations to translate into case-by-case trusting behavior is consistent with La Pierre's (1934) finding that generalized ethnic attitudes do not predict case-by-case discriminatory actions. It is also consistent with findings about the weakness of base-rate data (in this case, the gender stereotypes) by comparison with individuated data in the making of discriminations (Fischhoff & Bar-Hillel, 1984; Kahneman & Tversky, 1973; Locksley, Hepburn, & Ortiz, 1982; Lyon & Slovic, 1976).

In a somewhat different vein, the data provide a further demonstration of the ease with which "category-based expectancies" are overwhelmed by "target-based expectancies." As Jones and McGillis (1976) define them, the former derive from the judge's knowledge that a target person is a member of a particular class, category, or reference group, while the latter derive from prior information about the particular individual actor. Our subjects had only scant information about potential partners — what could be gathered absent any discussion, presumably by simple scrutiny — but that seems to have been enough to overcome the category-based expectancy that women will cooperate more than men.

There is a reservation. Our subjects were not exclusively students, but about 87% of them were. (Most of the remainder were unemployed townspeople.) Had we asked them about "*students* in general" as opposed to "*people* in general" it is possible we would have found a smaller difference by gender in generalized expectations; students might believe that cooperativeness is more gender-based in the nonstudent population than in the student one. If so, then their failure to distinguish by gender in their case-by-case responses would not be inconsistent with the relevant categories. Clearly, more analysis with more finely defined categories will be useful.

With Brown-Kruse and Hummels (1993), we point out that simply asking people to predict their own and others' behavior is likely to produce substantially different findings from observing actual behavior — a discrepancy that is likely, also, to be particularly great when the behavior in question, like cooperation, is ethically loaded. We speculate that scholars who identify essentialist differences between male and female moral behavior have tapped more into the gender-based stereotype than into what actually happens.

But what accounts for the generalized expectation that women will cooperate more than men? We resist the idea that such a firmly held belief — one held equally by male and female — has no basis in reality outside the laboratory; our subjects' perceptions, we feel, ought to be taken seriously.

Findings by Eagley and Steffen (1984, 1986) are suggestive. They show that gender-based stereotypes — in particular, the stereotype that women are more communal than men — are based on sex differences in the distribution of men and women across social roles. However much recent history has softened traditional patterns, women *do* occupy "helping" roles (caregiver to children, secretary, nurse, etc.) more than men. Consequently, it may not be an error to expect women outside the laboratory to cooperate more than men. It is simply a fact of life and — we propose — our subjects' generalized expectations simply qualify them as reasonably accurate social observers.

Granted that our subjects viewed each other as members of the categories "men in general" and "women in general," their failure to base their case-by-case expectations on gender is quite rational. *Within* roles we should expect no differences between genders in the frequency of cooperation — and neither should we expect such differences in the laboratory where subjects confront other individuals in situations that are in no way role-related. By this argument, our subjects' trust-related behavior reveals them, again, as reasonably accurate social observers.

Why women tend to occupy more "cooperative" roles is, of course, an important issue, but our data do not support any essentialist argument — for example, that women select themselves into such roles because of their cooperative natures. When women are "randomized out of their social roles," their cooperation rates are indistinguishable from those of men. Socialization, custom, persuasion, economic necessity, or a power imbalance between the genders (or some combination of them) might all explain role occupancy, but not — from our data — essentialist differences in cooperative dispositions.

The failure of essentialist arguments also suggests that the success of all-female groups depends on the ability of members to organize themselves for cooperative action — no more and no less than is the case for all-male or mixed-gender groups. (For a case study, see Schwartz-Shea & Burrington, 1990.) Perhaps all-female groups might turn the expectation of female cooperativeness to good use, but the organizational problems such groups confront are not fundamentally different in this respect from the problems confronted by all groups.

What criteria were subjects using to decide between trusting and not trusting others in our experiments? Orbell and Dawes (1991, 1993) propose that some players, at least, project from their own cooperate versus defect intentions to those of potential partners (Riker's "introspective" method), and it is possible that subjects who refused play with any of their five potential partners based that choice on the dominance of defection over cooperation (Riker's "utilitarian" method)[7]

Other target characteristics like age, race, and ethnicity are no less visible than gender, and might be a basis for trust — or mistrust — even in the laboratory where other contextual cues are missing. Unfortunately, we are not in a position to test such hypotheses. Most of our subjects were students, thus we had only little variance by age. And, while there was some variation by race and ethnicity, there was not sufficient for systematic

[7] On the other hand, those who opted out for some but not for others would seem to have been making case-by-case discriminations of some kind.

study; the populations of Eugene and the University of Oregon are overwhelmingly white.

We speculate, however, that findings about the relationship between such other target characteristics and trust will reflect our current findings about gender. Our experience with categories of people is inevitably context based, and when we are asked to predict the behavior of categories in the abstract, it makes sense for us to base our predictions on the contexts from which our experience has been gathered.

Yet, as our gender data suggest, we can disentangle our case-by-case judgments from such "context-based knowledge" when the circumstances of those judgments makes it plain that we should. Undoubtedly, few natural circumstances extract people from their everyday circumstances as neatly as does a fully randomized laboratory experiment. But our subjects appear to have responded to the fact that this had happened by ignoring the context-based knowledge they brought with them about the relationship between gender and cooperation.

Given that this rejection of context-based knowledge can happen with respect to gender—about which, as our data suggest, such "knowledge" is very strong—we see no reason why it could not happen with respect to any of the other social categories by which we develop our expectations about each other.

APPENDIX

A. Dollar payoffs for the five matrices

a.	2,2	–7,5
	5,–7	–5,-5
b.	2,2	–7,5
	5,–7	–2,–2
c.	2,2	–2,3
	3,–2	–1,–1
d.	2,2	–2,4
	4,–2	–1,–1
e.	2,2	–2,3
	3,–2	–1,–1

B. Sequence of plays by partner and matrix

	PLAY 1	PLAY 2	PLAY 3	PLAY 4	PLAY 5
A PLAYS WITH	F(e)	E(d)	D(c)	C(b)	B(a)
B PLAYS WITH	A(a)	F(b)	E(e)	D(d)	C(c)
C PLAYS WITH	B(c)	A(b)	F(d)	E(a)	D(e)
D PLAYS WITH	C(e)	B(d)	A(c)	F(a)	E(b)
E PLAYS WITH	D(b)	C(a)	B(e)	A(d)	F(c)
F PLAYS WITH	E(c)	D(a)	C(d)	B(b)	A(e)

REFERENCES

Aranoff, D., & Tedeschi, J. (1968). Original stakes and behavior in the prisoner's dilemma game." *Psychonomic Science, 12,* 79-80.

Baier, A. (1985). What do women want in a moral theory? *Nous, 53,* 53-62.

Brown-Kruse, J., & Hummels, D. (1993). Gender effects in laboratory public goods contribution: Do individuals put their money where their mouth is? *Journal of Economic Behavior and Organization, 22,* 255-268.

Dawes, R., McTavish, J., & Shaklee, H. (1977). Behavior, communication and assumptions about other peoples' behavior in a commons dilemma situation. *Journal of Personality and Social Psychology, 35,* 1-11.

Eagly, A., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100,* 283-308.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46,* 735-754.

Eagly, A. H., & Steffen, V. J. (1986). Gender stereotypes, occupational roles, and beliefs about part-time employees. *Psychology of Women Quarterly, 10,* 252-262.

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory and Cognition, 12,* 402-410.

Gauthier, D. (1986). *Morals by agreement.* Oxford, England: Clarendon.

Gilligan, C. (1982). *In a different voice; Psychological theory and womens' development.* Cambridge, MA: Harvard University Press.

Hamilton, S. (1986). *A comparison of coping styles in male and female young adults.* Dissertation, Psychology Department, University of Oregon, Eugene.

Javine, D. (1986). *A gender comparison: Cooperation for the public good.* Unpublished doctoral dissertation, Clinical Psychology, University of Nevada at Reno.

Jones, B., Steele, M., Gahagan, J., & Tedeschi, J. (1968). Matrix values and cooperative behavior in the prisoner's dilemma game. *Journal of Personality and Social Psychology, 8,* 148-153.

Jones, E., & McGillis, D. (1976). Correspondent inferences and the attribution cube: A comparative reappraisal." In J. H. Harvey, W. J. Ickes, & R. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 389-420). Hillsdale, NJ: Erlbaum.

Kahn, A., Hottes, J., & Davis, W. (1971). Cooperation and optimal responding in the prisoner's dilemma game: Effects of sex and physical attractiveness. *Journal of Personality and Social Psychology, 17,* 267-279.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251.

La Pierre, R. T. (1934). Attitudes versus action. *Social Forces, 13,* 230-237.

Locksley, A. Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology, 18,* 23-42.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica, 40,* 287-298.

Meux, E. P. (1973). Concern for the common good in an N-person game. *Journal of Personality and Social Psychology, 28,* 414-418.

Orbell, J., & Dawes, R. (1991). A "Cognitive Miser" theory of cooperators' advantage. *American Political Science Review, 85,* 515-528.

Orbell, J., & Dawes, R. (1993). Social welfare, cooperators' advantage and the option of not playing the game. *American Sociological Review, 58,* 787-800.

Orbell, J., Dawes, R., & van de Kragt, A. (1990). The limits of multilateral promising. *Ethics, 100,* 616-627.

Orbell, J., Schwartz-Shea, P., & Simmons, R. (1984). Do cooperators exit more readily than defectors? *American Political Science Review, 78,* 147-162.

Pearson, R. W., Ross, M., & Dawes, R. M. (1992). Personal recall and the limits of retrospective questions in surveys. In J. M. Tanur (Ed.), *Questions about questions; inquiries into the cognitive bases of surveys* (pp. 65-94). New York: Russell Sage Foundation.

Rapoport, A. (1967). Optimal policies for the prisoner's dilemma. *Psychological Review, 74(2)* 136-148.

Rapoport, A., & Chammah, A. (1965). Sex differences in factors contributing to the level of cooperation in a prisoner's dilemma game. *Journal of Personality and Social Psychology, 2,* 831-838.

Riker, W. (1980). Political trust as rational choice." In L. Lewin & E. Vedung (Eds.), *Social choice: Addresses, essays, lectures* (pp. 1-24). Dordrecht, The Netherlands: D. Reidel.

Rose, C. (1992). Women and property: Gaining and losing ground. *Virginia Law Review, 78,* 421-459.

Schuessler, R. (1989). Exit threats and cooperation under anonymity. *Journal of Conflict Resolution, 33,* 728-749.

Schwartz-Shea, P., & Burrington, D. (1990). Free riding, alternative organization and cultural feminism: The case of Seneca womens' peace Camp. *Women and Politics, 10,* 1-37.

Sell, J., Griffith, W., & Wilson, R. (1991). *Are women more cooperative than men in social dilemmas?* Unpublished paper, Rice University.

Stockard, J., & Johnson, M. M. (1992). *Sex and gender in society* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Stockard, J., van de Kragt, A., & Dodge, P. (1988). Gender roles and behavior in social dilemmas: Are there sex differences in cooperation and its justification? *Social Psychology Quarterly, 51,* 154-163.

Tullock, G. (1985). Adam Smith and the prisoner's dilemma. *Quarterly Journal of Economics, C,* 1073-1081.

Vanberg, V., & Congleton, R. (1992). "Rationality, morality and exit." *American Political Science Review, 86,* 418-431.

Yamagishi, T. (1988). Exit from the group as an individualistic solution to the free rider problem in the United States and Japan." *Journal of Experimental Social Psychology, 24,* 530-542.