Do Participants and Observers Assess Intentions Differently During Bargaining and Conflict? An Experiment

Eric S. Dickson*

Abstract

Political actors in settings of bargaining and conflict often find themselves uncertain about the intentions of their counterparts. This paper explores the psychology of how intentions are assessed using a novel experimental design involving imperfect-information versions of the ultimatum and dictator bargaining games, two paradigmatic models of conflict. Subjects are randomly assigned to one of three roles – the traditional proposer and recipient roles in these games, and a novel impartial observer role. Recipients and observers are given identical, but ambiguous, information about proposers' offers, and make post-play assessments of proposers' intentions that are rewarded based on accuracy. When uncertainty is sufficiently high, recipients' assessments of proposers' intentions are significantly lower than observers' assessments in the ultimatum game, in stark contrast to Bayesian predictions, but there is no evidence of any difference in the dictator game. The results suggest that actors' perceptions can be directly affected by the set of strategic alternatives they possess, independent of access to information. One interpretation is that the power to accept or reject may prime individuals to be more critical or negative in forming assessments given ambiguous evidence than they otherwise would be. If correct, this interpretation has important implications for theories of bargaining and conflict, and for the design of institutions for conflict resolution.

*Assistant Professor, Department of Politics, New York University. Mailing address: 19 West 4th Street, Department of Politics, New York University, New York, NY 10012. Email: eric.dickson@nyu.edu. Fax: (212) 995-4184. I am grateful to Becky Morton, Mik Laver, and Andy Schotter for their comments and suggestions.

1 Introduction

Political actors often find themselves uncertain about the intentions of their counterparts. While actors can sometimes learn about others' intentions by observing their performance at some task, at other times the evidence offered by such observations can be ambiguous or prove difficult to interpret. One key reason for this is that, in a complicated world, actions may not map neatly onto consequences. For example, if a governing authority agrees to suppress terrorist activity within its territory, but then fails to do so, the failure may indicate latent sympathies for the terrorists' cause – or it may have been the result of unanticipated difficulties caused in part by bad luck. In many settings involving bargaining or conflict, the assessments that actors ultimately make about their counterparts' intentions can influence the outcome of negotiations, the ability of two opposed parties to come to an agreement, or even the prospects for war and peace.

Given this, an understanding of how actors interpret ambiguous evidence about their counterparts' intentions would seem a crucial component of theories of conflict, conflict resolution, and political bargaining more generally. How do actors make judgments about others' motivations when presented with ambiguous evidence about the actions their counterparts have taken? Do particular features of actors' *strategic* positions intrinsically color the way in which they form judgments? Do participants in conflictual interactions form beliefs about one another in systematically different ways than a disinterested observer would – even controlling for the information to which actors have access?

These questions are important ones – yet differing research traditions within political science suggest very different answers. Game-theoretic and rational-choice conceptions of conflict and bargaining assume that individuals, upon encountering new information, accurately and dispassionately update their prior beliefs about counterparts' "types" according to Bayes' Rule. In sharp contrast, behavioral research traditions within political science and social psychology hold that actors' assessments of others' intentions are subject to systematic cognitive biases and may be influenced by features of their social and situational roles independent of the information to which they have access.

Psychologists have long conceived of a dichotomy between "situational" and "dispositional" attributions (Heider 1958; Jones and Davis 1965; Kelley 1967); in this language, the governing authority described above may have failed to suppress terrorist activity because the circumstances made it im-

possible for it to do so (situational) or because the authority did not make a proper effort due to its true underlying sympathy with the extremists' cause (dispositional). It has long been argued that people have a systematic tendency to overweight the dispositional component, underweighting situational factors in assessing others' intentions, a form of bias referred to as the "fundamental attribution error" (Ross 1977) or "overattribution" (Jones 1979). Existing literature suggests however that the nature of actors' attributions may vary depending on various contextual factors, because the relative salience of dispositional and situational factors may vary for different kinds of actors or in different settings. A classic distinction in the literature involves the "actor-observer" difference (Jones and Nisbett 1971), by which actors' attributions of their *own* actions differ systematically from observers' attributions of those actions. Other researchers argue that the nature of bias in formulating attributions varies depending on, among other things, whether observed outcomes are positively- or negatively-valued (Van der Plight and Eiser 1983; Tillman and Carver 1980); whether the actor whose actions are being attributed is liked or disliked (Taylor and Koivumaki 1976); and whether that actor is of the same or a different social identity group than the evaluator (Chatman and von Hippel 2001). Nonetheless, the way in which, and the extent to which, individuals' judgments about others' intentions vary as a function of their positions in strategic environments remains a key, and largely open, question.

Within political science, international relations scholars have arguably given the most explicit attention to the possible presence and potential effects of cognitive biases in conflict and bargaining (e.g., Jervis 1976; Jervis, Lebow, and Stein 1985; Mercer 1996; see McDermott 2004 for a comprehensive review of the political psychology literature in international relations). For example, both Rosenberg and Wolfsfeld (1977) and Mercer (1996) use insights from the psychology literature to develop models of attribution behavior in settings of international conflict; in their models, actors' attributions are affected both by the desirability of the evaluated outcome and the extent of inherent sympathy with the actor whose intentions are being assessed. The methods employed by these scholars to illustrate and test their models reflect two key methodological strategies present in the literature: Rosenberg and Wolfsfeld employ survey methods, looking at differences in attributions made by members of different social identity groups (in their case, attributions by students of Arab and Jewish origin of actions taken in the Arab-Israeli conflict), while Mercer employs detailed case studies of historical episodes.

While the existing literature suggests the importance of non-informational factors in forming assessments of others' intentions, it does not do so without reservations or contradictions. For example, Jervis, in describing literature on the influence of desires and fears on perception, notes with concern

the fact that the relevant experiments typically lack incentives for accuracy (1976: 357), a critique that has long been leveled at prevailing methodologies in psychology research by experimental economists (see, e.g., Kagel and Roth 1995; Camerer 2003). Indeed, some results within the attribution literature itself highlight this concern; Tetlock (1985), for example, demonstrates that the "fundamental attribution error" can in some circumstances be mitigated by making subjects "accountable" for their judgments by having to justify them. Measurement issues pose major challenges for both traditional survey and case study designs as well; surveys eliciting assessments of complex real-world incidents generally cannot offer realistic incentives for accuracy, while the true private beliefs of historical actors as events unfolded in their time are also often elusive.

Other methodological concerns persist as well. Because the game-theoretic and social-psychological literatures have generally been motivated by different theoretical questions, typical existing studies do not pose or answer research questions in a way that offers a clear test of Bayesian models of belief formation under uncertainty. Gerber and Green (1999) as well as Taber and Lodge (2006) critique the related psychological literature on confirmatory biases along these lines; because the relevant studies typically do not measure or account for subjects' prior beliefs, or completely control for subjects' access to information, they are not well-placed genuinely to falsify Bayesian, rational accounts of the psychological phenomena they claim to demonstrate. Parallel critiques are also applicable to much existing political science survey and case study research. For example, a finding that members of different pre-existing social identity groups offer different judgments about a given event in a survey does not, in itself, demonstrate the existence of a "bias" in the sense of a violation of Bayes' Rule. Members of different social groups can, and probably do, experience novel events with different prior beliefs in tow, based on differing past experiences as well as exposure to different sources of information. All of these issues suggest that the inferences necessary to convincingly demonstrate non-Bayesian biases in judgments are difficult if not impossible to make in the kinds of observational data typically available in political science.

These features of the existing literature suggest the potential usefulness of a controlled laboratory experiment, in which individuals are randomly assigned to different roles, in which individual access to information can be controlled for carefully, and in which individual assessments of others' intentions can be elicited in a way that offers an incentive for accuracy. In tandem, random assignment and the ability to control for information allow for clean causal inferences about the effects of individuals' roles on the way in which they formulate attributions – causal inferences that are difficult if not impossible

to establish with confidence in observational data. Further, a controlled laboratory environment allows for direct manipulation of features of the bargaining or conflict setting across different experimental treatments. Measurement of any across-treatment variation in biases that may exist can potentially offer novel insights into the psychological mechanisms underlying these biases, greatly aiding the project of systematic theory building. While the interpretation of laboratory experimental results invariably involves questions of external validity, it is arguably a "tough test" of hypothesized biases in judgments to attempt to measure them in a stylized, controlled environment. If individuals exhibit robust patterns of bias, even when they have been *randomly* assigned to strategic roles in a simple experimental scenario, it seems reasonable to hypothesize that such biases might also be associated with actors in analogous roles in more complex, naturalistic, and emotionally-charged settings of bargaining and conflict.

This paper describes a laboratory experiment on the assessment of others' intentions during bargaining and conflict that meets all of the criteria specified above. The experimental protocol employs imperfect-information variations of the ultimatum game (Gueth, Schmittberger, and Schwarz 1982) and the dictator game (Forsythe et al 1994), two interactive scenarios that are often taken to be paradigmatic models of bargaining and conflict in political science scholarship. The ultimatum game consists of a simple model of bargaining in which a "proposer" makes an "offer" that is subsequently accepted or rejected by a "recipient" – upon acceptance, both parties receive the benefits detailed in the offer, whereas upon rejection, both parties receive nothing. The dictator game, in contrast, describes a setting in which a "proposer" is free to impose her "offer" on a "recipient" whether or not that recipient extends his approval.[1] The basic elements of these simple scenarios are to be found at the core of many bargaining and conflict processes. Half of the experimental sessions described in the paper are devoted to the ultimatum game, while the other half are devoted to the dictator game.

The experimental design adapts the ultimatum and dictator games to the study of intention attribution by varying the classic context of the games in three key ways. First, in both games, subjects are randomly assigned into one of *three* strategic roles – the proposer and recipient roles described above, as well as the novel role of "observer." Second, in both games, recipients and observers are given identical, *imperfect* information about the offer that was made by the proposer. And third, after the play of either game, recipients and observers are asked to use the ambiguous information they were given to formulate an assessment of what they think the proposer's offer actually was – and are

[1]For clarity of prose, throughout the paper proposers are referred to using female pronouns, while recipients are referred to using male pronouns.

paid depending on the accuracy of these assessments. Because individuals are randomly assigned to strategic roles, recipients and observers should share the same distribution of prior beliefs; and because recipients and observers also receive identical information about proposers' offers, the Bayesian, rational framework requires that recipients' and observers' assessments should be the same on average. As such, the experimental design is well-suited to a sharp test of Bayesian, rationalistic notions of how actors make judgments about others' intentions.[2]

The experimental results suggest that the Bayesian predictions can be soundly rejected in some circumstances, but not in others. Data from the ultimatum game sessions exhibits a striking pattern by which recipients make lower assessments of proposers' offers than observers do when the level of uncertainty about the offers is sufficiently high – and that this gap is substantial when offers appear more likely to have been relatively poor, but shrinks to statistical insignificance when offers appear more likely to have been relatively generous. In contrast, data from the dictator game sessions is quite consistent with the Bayesian prediction that there should be no gap apparent between recipients' and observers' assessments of proposers' intentions. Recipients and observers differ in two key ways in the ultimatum game – the outcome is payoff-relevant for recipients (but not observers), and recipients have an opportunity to respond to offers while observers are passive. Considering the ultimatum game in isolation, it is psychologically plausible that either of these two differences might induce recipients and observers to make different attributions. In the dictator game, however, they differ only in the former respect. The fact that no difference in assessments across roles is observed in the dictator game – despite the fact that the distributions of offers made in the two games are quite similar, and despite the fact that the overall pools of assessments made in the two games are comparably heterogeneous – suggests, along with other controls built into the experimental design, that recipients' assessments may be lower than observers' are in the ultimatum game *because* recipients have the opportunity to reply to proposers' offers, whereas observers do not.[3] This result is consistent with a psychological mechanism

[2]Note that the participant-observer distinction studied here is distinct from the "actor-observer" studies cited above. This paper investigates the potential effects of two actors' distinct strategic positions on their assessments of a third actor's intentions; the "actor-observer" paradigm in the psychology literature compares an individual's attributions of his own actions to disposition or to circumstances, relative to the attributions made by an observer of that individual.

[3]Section 4.4 argues that this account is more consistent with the data than several other alternative explanations.

by which the strategic power to accept or reject an offer may *prime* an individual to be more critical or negative than the same individual would be in the absence of such power. This interpretation suggests that individuals' strategic alternatives may have *in themselves* a direct impact on attributions and perception, independent of the information that individuals actually receive about others' behavior. If true, this account, novel in the literature, has profound implications for our understanding of the psychology of bargaining and conflict, as well as the optimal design of institutions for bargaining and conflict resolution.

The remainder of the paper is organized as follows. Section 2 describes the experimental procedures. Sections 3 and 4 describe the experimental results. Section 3 presents data describing behavior in the ultimatum and dictator games themselves, while Section 4 analyzes subjects' responses to the post-play question eliciting their assessments of the proposers' intentions. Section 5 discusses potential implications of the results, suggests directions for future work, and concludes.

2 Experimental Procedures

The experiments were carried out in an experimental social science lab at a large university. The experimental results come from data collected in eight experimental sessions involving a total of 156 subjects. Subjects signed up for the experiment via a web-based recruitment system that draws from a broad pool of potential participants; individuals in the subject pool are mostly undergraduates from around the university, though a smaller number came from the broader community. Subjects were not recruited from the author's courses, and all subjects gave informed consent according to standard human subjects protocols. Subjects interacted anonymously via networked computers; the experiments were programmed and conducted with the software z-Tree (Fischbacher 1999).

Of the eight experimental sessions, four (involving 78 subjects) were devoted to an imperfect-information version of the ultimatum game, and four (also involving 78 subjects) were devoted to an imperfect-information version of the dictator game. These sessions are summarized in Table 1.

TABLE 1 ABOUT HERE

In every experimental session, the number of subjects taking part was equal to a multiple of three. Each of the subjects present at a given session was randomly assigned by the experimental software to one of three roles. In the experimental instructions and protocol, these roles were given the neutral labels of "Role 1," "Role 2," and "Role 3." An equal number of subjects was assigned to each of these

roles, and the role assignments remained fixed through the duration of an experimental session, so that each individual subject had experience of only one role. In both the ultimatum and dictator game variations, subjects in "Role 1" acted as "proposers"; subjects in "Role 2" acted as "recipients"; and subjects in "Role 3" served as "observers."[4]

Once roles had been assigned, a set of instructions describing the structure of the experimental game was read aloud from a text that was also distributed in hard copy form to all subjects, regardless of their specific assigned roles. Such a procedure is generally employed in laboratory studies of behavior in games as a way of promoting understanding of lab scenarios and inducing common knowledge. In the present experiment, it also aids in strengthening the inferences that can be drawn from comparing the assessments formulated by recipients and observers, by ensuring that subjects in both of these roles are informed about the experiment in precisely the same way. It is also worth noting that the instructions for the ultimatum and dictator games differed only in the most minimal respects necessary.[5] After completing a brief on-screen quiz about the instructions, administered both to measure and to promote understanding, subjects participated in a number of periods of the particular game to which their session was devoted.[6]

Each period consisted of two parts: first, a play of the relevant experimental game, and second, a post-play "question" in which recipients (Role 2) and observers (Role 3) had an opportunity to increase their earnings by giving their best guesses about the proposer's (Role 1) choice. Subjects' payoffs both for the games and for the post-play questions were demarcated in experimental "tokens" that were subsequently converted into currency at the rate of 100 tokens to US$1. Subjects' total earnings consisted of the sum of their payoffs in each period, plus a showup fee of US$7.

2.1 The Experimental Games

At the beginning of each period, subjects were randomly rematched into a new "group" consisting of one proposer, one recipient, and one observer. Both the ultimatum and dictator games began with each group's proposer (Role 1) choosing a potential allocation of 100 tokens, offering x tokens to the

[4]Comparisons between recipients' and observers' assessments of proposers' intentions were therefore carried out within the context of a between-subjects experimental design.

[5]The text of all instructions to subjects can be found in the Referees' Supplemental Appendix (which will also be posted online at the time of publication).

[6]The number of periods in a given session was not announced to subjects in advance.

recipient (Role 2) in her group while proposing to keep $100 - x$ tokens for herself. Proposers were allowed to choose any integer value of x between 0 and 100 (inclusive).

A key novelty of the ultimatum and dictator game settings in this experiment is that recipients (and observers) were not told the value of the offer x.[7] Instead, for a particular offer x, the recipient and the observer in the proposer's group were given identical values x_{min} and x_{max} and told simply that $x_{min} \leq x \leq x_{max}$ – that is, that x fell within the "range" $[x_{min}, x_{max}]$. The instructions to subjects did not contain any specific information about the ways in which the "range" was generated from any particular offer x. Ultimately, this feature of the experimental design allows for comparisons of how participants (recipients) and observers assess proposers' intentions when, as in most real-world settings, the underlying structure of uncertainty is not generally understood.[8] The "widths" of these "ranges" (e.g., the value $x_{width} = |x_{max} - x_{min}|$), which always took on a value from the set $\{15, 20, 25, 30, 35, 40\}$, were randomly generated, as was the location of the range midpoint $\frac{x_{min}+x_{max}}{2}$ relative to x, subject to a constraint that $x_{min} \geq 0$ and $x_{max} \leq 100$ (so that the ranges never included offers that were infeasible for the proposer to make).[9] Both x_{min} and x_{max} were always equal to a multiple of 5.

[7]While other studies have explored behavior in the ultimatum game under one-sided imperfect information (e.g., Mitzkewitz and Nagel 1993; Rapoport and Sundali 1996; Rapoport, Sundali, and Seale 1996), these studies operationalized uncertainty differently, focused on different research questions, and did not employ an observer role.

[8]This feature of the design does not allow for tests of whether *any specific subject's* assessments are Bayesian – individual subjects' prior beliefs about the proposers' inclinations and the uncertainty-generating process are unobserved. However, the design *does* offer a clean test of a key, necessary consequence of the Bayesian view – that, because subjects are randomly assigned to the recipient and observer roles, their assessments should on average be the same, given the same information about proposers' offers.

[9]x_{min} and x_{max} were actually generated for a given offer x in the following way. First, a preliminary range half-width r was randomly drawn from a uniform distribution on $[7.5, 17.5]$. A preliminary range midpoint m was then drawn from a uniform distribution on $[x - r, x + r]$. Preliminary values x'_{min} $[x'_{max}]$ were then calculated by rounding $m - r$ $[m + r]$ down [up] to the nearest multiple of 5. Finally, if $0 \leq x'_{min} < x'_{max} \leq 100$, x_{min} and x_{max} were set equal to x'_{min} and x'_{max} respectively; if $x'_{min} < 0$, x_{min} was set to 0 and x_{max} to $x'_{max} - x'_{min}$; and if $x'_{max} > 100$, x_{max} was set to 100 and x_{min} to $x'_{min} - (x'_{max} - 100)$.

In the ultimatum game sessions, once x_{min} and x_{max} were known both to recipient and observer, recipients chose to accept or to reject the offer, based on this imperfect information. If a recipient accepted an offer whose *actual* value had been x, he received a payoff of x tokens for the game, while the proposer received $100 - x$ tokens. If on the other hand a recipient rejected an offer whose actual value had been x, both recipient and proposer received 0 tokens.

In the dictator game sessions, on the other hand, both recipient and observer received information about the offer in the form of x_{min} and x_{max}, but the recipient did not have the ability to accept or reject offers. Instead, each proposer's allocation was automatically adopted, leaving the recipient with x tokens and the proposer with $100 - x$ tokens as game payoffs. Thus, proposers' offers are relevant to recipients' payoffs in both the ultimatum and dictator games, but recipients have an ability to react to these offers only in the ultimatum game. This aspect of difference between the ultimatum and dictator games, along with their other structural similarities, will be useful in interpreting patterns of recipient-observer differences in the two games.

Observers (Role 3) did not make choices in either the ultimatum or dictator game, and their payoffs were not affected by the outcomes of the games. Observers simply received a flat payoff of 40 tokens as a game payoff in each period of play.

2.2 Post-Play Assessments: Recipients' and Observers' Interpretations of Proposers' Intentions

Once play of the ultimatum or dictator game in a given period was complete, subjects proceeded to the second part of the period, involving a post-play "question" that offered subjects an additional opportunity to earn payoffs. It is important to note that subjects received no feedback about the outcome of game play before proceeding to the post-play questions.[10]

In the post-play question phase, recipients (Role 2) and observers (Role 3) were asked to provide their best guess as to the value of x – that is, the amount that had actually been offered to the recipient by the proposer.[11] As a way of motivating subjects to pay attention and make the most accurate assessments possible, recipients and observers were paid for guesses that were sufficiently accurate. Specifically, if a recipient or an observer entered a guess of y when the actual offer had been

[10]More details about experimental feedback to subjects are contained in Section 2.3.

[11]The text associated with this question was identical for recipients and observers and can be found in the reviewers' appendix.

x, he received a payoff (in tokens) of $max(100-10|x-y|,0)$. That is, a correct guess yielded a payoff of 100 tokens; a guess that was incorrect but which deviated from the true answer by no more than nine "units" received a payoff of 100 tokens *minus* ten tokens for each "unit" between their guess and the true value of x; and any incorrect guess deviating from the true answer by 10 or more "units" received a payoff of 0 tokens.[12]

The assessments y that recipients and observers make about proposers' offers are the key objects of interest in the experimental design. In particular, the central research question involves any differences that may exist in how recipients and observers perceive identical information about proposers' intended behavior, and how such differences may vary across different strategic settings. Several aspects of the experimental design strengthen the inferences that can be drawn from comparing recipients' and observers' assessments y. After a proposer decision in either game, the recipient and the observer receive the *same* information about x in the form of x_{min} and x_{max}. This control for access to information, along with the random assignment of subjects to the roles of recipient and observer, allows for the possibility of causal inference about the effects of an individual's strategic role (e.g., recipient vs observer) on that individual's perceptions of proposers' intentions, as measured by the assessments y.

2.3 Experimental Feedback

The previous two subsections described the structure of subjects' choices in a given period of play. As the experimental sessions consisted of multiple periods, it is important to describe the feedback received by subjects from period to period in order to account for the possibility of learning by subjects and any time trends that might appear in the data. The feedback that subjects received in the experiment was quite limited. Subjects were not given any information as to their payoffs, either for the play of the games or for the post-play questions, until after their participation in the experiment was complete. In particular, recipients and observers did not receive any direct feedback as to what the true value

[12]While recipients and observers were answering this question, proposers (Role 1) were given the values x_{min} and x_{max} that had been generated from their offer x, and informed of the question that had been posed to recipients. They were then asked to provide their best guess as to what guess y would be entered by the recipient in her group. Labelling the proposer's guess z, the proposer was paid $max(50-5|y-z|,0)$ tokens for her answer. Information about z is not relevant to the research question explored here, so this aspect of the experimental design will not be discussed further.

of x had been, either through their payoffs in the game (recipients) or through the accuracy of their guesses (both recipients and observers). Proposers were similarly not told whether their offers had been accepted or rejected by recipients. While subjects did not receive direct information about their payoffs, subjects did observe values of x_{min} and x_{max} during every period. These observations made it unavoidable that recipients and observers did receive some *indirect* feedback about the distribution of proposers' offers, and that subjects in all roles received indirect feedback about the way in which ranges were generated.

As motivated above, the experiment was designed to explore any causal effect of actors' strategic roles on perception formation in the context of a setting of ambiguity about how agents' actions map onto observable consequences. Such a setting is not only well-suited to the research question, but it also heightens the prospects for external validity from a psychological standpoint. The decision to offer subjects only limited feedback in the research design was intended to further these objectives. As recipients and observers learn more about the distribution of proposers' behavior, and learn more about the structure of uncertainty in the experimental scenario, any inherent tendency towards differences in participant and observer cognition is increasingly likely to be washed out.

3 Experimental Results: Game Behavior

This section briefly describes subject behavior in the imperfect-information ultimatum and dictator games themselves. Analysis of recipients' and observers' post-play assessments of proposers' intentions is contained in section 4.

3.1 Behavior in the Imperfect-Information Ultimatum Game

TABLE 2 ABOUT HERE

The left half of Table 2 presents the distribution of offers x made by proposers during all periods of the four ultimatum game sessions. Overall, the mean offer x over all periods of the four ultimatum game sessions shown in these columns was 26.1 (standard deviation = 20.9). The data suggest a tendency for offers to decrease modestly over time; the mean offer x during the first eight periods was 29.5 (standard deviation = 21.2), while the mean offer x during the last seven periods was 22.2 (standard deviation = 19.8). Proposers' offers in this imperfect-information version of the ultimatum game are somewhat lower on average than the offers typically observed in standard ultimatum game experiments that do

not involve imperfect information; a meta-analysis of numerous existing studies (Camerer 2003, p. 49) suggests that mean offers of between 30-40 percent of the amount available for distribution are more typical.[13]

Nonetheless, a striking feature of the left half of Table 2 is the high degree of heterogeneity in proposer behavior. While the modal offer was 0 (48 times, or 12.3% of all offers), the second- and third-most common offers were 40 (39 times, 10.0%) and 50 (also 39 times, 10.0%). Many other offers took on intermediate values spread broadly between 0 and 50; fully 95.1% of the offers were less than or equal to 50. The heterogeneity in proposer behavior, along with random re-matching of subjects into new groups in every period, helped to ensure that recipients and observers were shown substantially varying ranges in x across periods. For example, 38.7% of the ranges revealed to recipients and observers had 0 as the lower bound, while 40.0% of the revealed ranges included 50. The presence of such variation is an important precondition for the experiment's aim of measuring differences, if any, between recipient and observer perceptions of proposers' behavior in the context of the repeated-trials experimental design. In an environment characterized by minimal variation, the distribution of proposers' behavior might well become obvious right away, reducing recipients' and observers' uncertainty to very low levels, and thereby minimizing the potential for uncovering any systematic differences that might exist in how participants and observers tend to form beliefs.

The behavior of recipients in the imperfect-information ultimatum game sessions suggests that, as in standard ultimatum game experiments, recipients are willing to reject offers on grounds of fairness – even though such rejections can be costly, and even though subjects are randomly rematched into new groups from one period to the next. In the aggregate, 64 of the 390 offers made during the four ultimatum game sessions were rejected. This 16.4% rejection rate is within the range observed in the literature on standard ultimatum game experiments (Camerer 2003, pp. 53-55). The data indicate that a key factor in recipients' decisions to reject offers is whether or not $x_{min} = 0$. When $x_{min} = 0$, the rejection rate is quite large, at 36.4% (55 rejections out of 151 offers). When x_{min} takes on a greater value – that is, when the range reported to recipients and observers does not include zero – rejections are in contrast extremely rare, occurring only 3.8% of the time (9 rejections out of 239 offers).

[13]This finding, consistent with Rapoport, Sundali, and Seale (1996) as well as other imperfect-information studies, naturally suggests that proposers exploit recipients' uncertainty in choosing the offers they make.

3.2 Behavior in the Imperfect-Information Dictator Game

The distribution of offers x made by proposers during all periods of the four dictator game sessions is presented in the right half of Table 2. Overall, the mean offer x over all periods of the four dictator game sessions was 23.0 (standard deviation = 22.3), strikingly similar to the mean offer value in the ultimatum game sessions. The mean offer x during the first eight periods was 24.4 (standard deviation = 21.8), while the mean offer x during the last seven periods was 21.3 (standard deviation = 22.9), suggesting, as in the ultimatum game, a modest tendency towards decreasing offers over the course of the experiment. These values are quite typical of proposers' offers in standard dictator game experiments, which typically average about 20 percent of the amount available for distribution (Camerer 2003, pp. 56-58), suggesting that proposers' offers were not greatly affected by the imperfect-information nature of the dictator game treatments.

Further, as in the ultimatum game sessions, there is a considerable degree of heterogeneity in proposers' offers in the dictator game. The modal offer in the dictator game was again 0 (77 times, or 21.1% of all offers), while the second-most-common offer was again 50 (45 times, 12.3%). Many other offers took on values spread broadly between 0 and 50; fully 94.3% of the offers were less than or equal to 50. This diverse set of offers was reflected by a diverse set of ranges revealed to recipients and observers; 49.9% of the announced ranges had 0 as the lower bound, while 27.9% of the announced ranges included 50.

Taken together, these descriptive statistics suggest a distribution of proposer behavior in the dictator game that is strikingly similar to the distribution of proposers' ultimatum game offers, in terms not only of mean values, but also of other distributional properties including the modal values and the overall level of variance (heterogeneity in offers). These similarities between *proposer* behavior in the ultimatum and dictator games offer a set of useful "controls" in studying any differences *between* games in how recipients and observers formulate assessments compared to one another. Because the distributions of offers these actors observe are so similar in the two games, any differences in recipient-observer differences across the games can arguably be related to the psychology of subjects' responses to the structures of the games themselves.

4 Experimental Results: Post-Play Assessments

4.1 Comparing Recipients and Observers

The core empirical strategy for comparing recipients' and observers' assessments of the proposers' actions in the ultimatum and dictator games employs regression analysis. The basic regression specification, employed separately for the ultimatum game and for the dictator game, is:

$$ASSESS_i = \beta_0 + \beta_1 RANGEMEAN_i + \beta_2 OBSERVER_i + \beta_3 RANGEMEAN_i * OBSERVER_i + \epsilon_i$$

Any given play of either game contributes two data points to the corresponding regression – one corresponding to the recipient's assessment of the size of the offer, and one corresponding to the observer's assessment. These assessments of x are represented by the dependent variable ASSESS in the regression equation. The first independent variable, RANGEMEAN, corresponds to the midpoint (mean) of the possible range of offers communicated to the recipient and the observer, that is, $x_{mean} = \frac{x_{min}+x_{max}}{2}$. If subjects respond at all to the information described by the ranges, a positive relationship between ASSESS and RANGEMEAN would be expected. The second dependent variable, OBSERVER, is a dummy variable equal to 1 for subjects in the observer role and 0 for subjects in the recipient role. This term allows the estimation of a direct effect of a subject's strategic role on belief formation. Finally, the third dependent variable is an interaction term, RANGEMEAN*OBSERVER, which allows the effect of OBSERVER to vary depending on the value of RANGEMEAN. For example, it allows the effect of OBSERVER to be different for offers more likely to be poor (those with a lower value of RANGEMEAN) than it is for offers more likely to be good (those with a higher value of RANGEMEAN). All of the regression analyses in the paper were carried out using ordinary least squares (OLS), with standard errors clustered at the level of individual subjects. Note that the research design implies that subjects' exposures to different communicated ranges of offers is balanced between recipients and observers at the aggregate level, since in each play of either game, any given range is communicated to one recipient and to one observer.

This regression specification offers an estimation strategy that is simple, but that nonetheless goes to the heart of the research question. However, this specification does not take into account two structural factors of the experimental setting that may affect the way in which participants and observers formulate assessments: the varying degree of ambiguity in the information provided to the subjects (i.e., varying x_{width}), and the fact that the experimental sessions are carried out over a number of periods.

The central question motivating this research concerns the extent to which individuals' strategic

roles may affect the way in which they come to perceive others' intentions. If individuals' roles do causally affect their perceptions, controlling for the information to which they have access, it stands to reason that the nature and magnitude of such effects will depend on what *kind* of information they are exposed to. Information that conveys another actor's intentions with little ambiguity and little room for alternative interpretations may be unlikely to lead different kinds of actors to differing conclusions, even if more ambiguous information may have the potential to do so. The choice to expose subjects to differing degrees of ambiguity in the form of range widths varying between 15 and 40 was influenced by the lack of guidance in the literature as to "how much" ambiguity, in a quantitative sense, might be required to observe any role differences that may exist. The first analysis carried out below estimates the basic regression specification separately for ultimatum-game data corresponding to each of the range width values. These results, along with follow-up analyses, suggest that it is appropriate to split the data sample into two halves, corresponding to settings of "Low Uncertainty" (range widths of 15, 20, and 25) and settings of "High Uncertainty" (range widths of 30, 35, and 40), because the estimations pool statistically within these categories but not across them. Thereafter, recipient and observer assessments are analyzed under each of these two regimes separately.

A related issue is that the experimental sessions unfold over time. While recipients and observers received only indirect feedback as described above, nonetheless they did have some opportunity to learn about the nature of uncertainty in the experiment (and the distribution of proposers' behavior) from period to period. Such learning would of course decrease subjects' *effective* level of uncertainty over the course of an experimental session; therefore in the context of this study it is to be expected that any recipient-observer differences would likely decrease over time.[14] The analyses below track this

[14]The experiment is meant to explore whether observers and conflict participants make systematically different assessments of a counterpart's intentions within the ambiguous context of a novel event or crisis. The ideal experimental design given this question would arguably require a very large number of subjects to make one assessment only; this was not feasible practically, so a smaller number of subjects made assessments over a number of periods instead, with random rematching between periods. Given subjects' repeated exposure to the *same* uncertainty environment, time trends mitigating any treatment effects are only to be expected. However, this by no means implies that participant-observer differences should be expected to wash out in actual *ongoing* conflicts, in which successive events or crises are likely to be uncertain in different ways and pose novel inferential problems. Any gap between participant and observer beliefs about an agent after *one* novel event will translate into differing *prior* beliefs

intuition first using interaction terms that allow for time trends in recipient-observer differences, but also using split-sample analyses and graphical depictions of how behavior evolves period-by-period.

Finally, in comparing recipients' and observers' assessments of proposers' intentions, it is crucial to ensure that any pattern of differences that is discovered is the result of a robust psychological phenomenon rather than being driven by a small number of outlying data points. This issue is particularly important given the between-subjects experimental design. If a given subject fails to engage or understand the experimental tasks, then he or she may make nonsensical, outlying choices that could potentially induce differences in the data between recipients and observers where no meaningful difference exists. Perhaps the most intuitive method of scanning the data with such concerns in mind is to flag any assessments about the offer x that do not fall within the range announced to recipients and observers (that is, assessments $y < x_{min}$ or $y > x_{max}$). Such "out-of-range" assessments are likely to have been the result of misunderstanding, misreading, or a typographical error by a subject in entering an assessment; "out-of-range" assessments not only correspond to infeasible values of the offer x, but from a payoff standpoint, are weakly dominated choices as well.[15] In the ultimatum game sessions, subjects made a total of only 17 "out-of-range" assessments (out of 390, or 4.4%). Two subjects made "out-of-range" assessments in multiple periods (5 and 10 times, respectively), suggesting that these subjects either failed to understand the experimental task or were motivated by factors other than accuracy in reporting assessments. In the dictator game sessions, subjects made a total of only 5 "out-of-range" assessments (out of 365, or 1.4%); no subject made such assessments in multiple periods.

Tables 3-6b, below, describe a series of regression analyses carried out using a dataset from which all "out-of-range" assessments, as well as *all* assessments from the two "repeat-offender" ultimatum-game subjects described above, have been dropped. Such analyses are arguably appropriate because, for the reasons given above, the dropped assessments are likely either to have been entered mistakenly, or to have been entered by subjects who did not understand the most basic features of the experimental setting. Further, because both "repeat-offender" subjects happened to be randomly assigned to the

about that *same* agent before the *next* novel event – so that differences in perception may in fact be self-reinforcing through ongoing conflicts. For a first experimental study in this area, though, it seemed imprudent to vary the uncertainty structure dramatically from period to period before understanding behavior in a simpler context, so this remains a question for future research.

[15] Assessments $y < x_{min}$ are weakly dominated by $y = x_{min}$; assessments $y > x_{max}$ are weakly dominated by $y = x_{max}$.

observer rather than the recipient role, the potential exists for these two subjects' outlying behavior to induce recipient-observer differences where no meaningful differences exist. Because experimental data points should never be dropped lightly, it is worth noting explicitly that the exclusions described above actually bias *against* the main experimental results. Because the (Bayesian) null hypothesis is that there should be no difference between recipients' and observers' assessments, dropping outlying points that increase estimates of such differences is arguably the more conservative strategy in attempting to demonstrate and estimate such differences.[16]

4.2 Recipients' and Observers' Assessments Differ in the Ultimatum Game

TABLE 3 ABOUT HERE

As noted above, from a theoretical perspective it is natural to suppose that the magnitude of any differences between recipients' and observers' assessments may vary depending on the degree of uncertainty present in a given situation. Table 3 therefore begins the analysis by carrying out the basic regression specification separately for each sub-sample corresponding to one possible value of $x_{width} = |x_{max} - x_{min}|$ – that is, $\{15, 20, 25, 30, 35, 40\}$. Keeping in mind that estimations within these sub-samples are imprecise because of the limited amount of data within each, a clear pattern nonetheless emerges. The estimated coefficients for OBSERVER are of varying sign and small magnitude for the lower values of $x_{width} = \{15, 20, 25\}$, but the estimated coefficients are of uniformly positive sign and of much larger, roughly comparable magnitudes for the higher values of $x_{width} = \{30, 35, 40\}$, offering a preliminary suggestion that observers may make higher assessments of proposers' intentions than recipients do under these conditions. Table 3 therefore suggests a natural split-sample estimation strategy, in which the data is divided into "Low Uncertainty" ($x_{width} = \{15, 20, 25\}$) and "High Uncertainty" ($x_{width} = \{30, 35, 40\}$) situations.[17]

[16]TO EDITORS AND REFEREES: The parallel full-dataset regression tables are contained in Tables 3R, 4aR, 4bR, 5R, 6aR, and 6bR of the Referees' Supplemental Appendix. They can be moved into the paper if this is your collective preference.

[17]Regressions incorporating interactions between x_{width} value dummies and the variables OBSERVER and RANGEMEAN*OBSERVER indicate that the effects of these variables are not statistically different for different values of x_{width} *within* the Low and High Uncertainty categories, but are different for values of x_{width} *across* the categories, offering further justification for the split-sample approach. While it was expected that the magnitude of any recipient-observer differences would vary with the extent of

TABLES 4a,4b ABOUT HERE

The results of this split-sample strategy are displayed in Tables 4a and 4b. Column (1) in both Tables carries out the same basic estimation as in Table 3, pooling the data within each uncertainty category. As foreshadowed, the results suggest no statistically significant differences between observers and recipients in Low Uncertainty ultimatum games, but a striking relationship between subjects' randomly-assigned strategic roles and their assessments of proposers' intentions in High Uncertainty ultimatum games. Across all periods of the ultimatum game sessions, the positive coefficient for OBSERVER indicates that, *ceteris paribus*, observers' assessments of proposers' offers are significantly higher than recipients' assessments are ($p < 0.05$) for low values of RANGEMEAN. The marginally significant negative coefficient for the interaction term RANGEMEAN*OBSERVER ($p < 0.08$), in contrast, offers an indication that this gap between observers' and recipients' assessments may *decrease* as RANGEMEAN *increases* – that is, as proposers' offers appear to be more generous.

As described above, intuitions that subjects may learn over time about the structure of uncertainty in the experiment, thereby reducing the *effective* level of uncertainty, call for analyses in which any evolution in recipient-observer differences can be discerned. Tables 4a and 4b approach this question in two distinct ways. Column (2) of both tables details a specification allowing the effects of OBSERVER and RANGEMEAN*OBSERVER to vary over TIME, using interaction terms (where TIME is simply the period number minus one). The overall marginal effect of OBSERVER remains statistically insignificant over time under Low Uncertainty (Table 4a).[18] However, a strongly significant,

uncertainty, why the data should partition so neatly into these two categories is not entirely clear. One possibility is that any recipient-observer differences in the Low Uncertainty setting are blurred out by the assessment elicitation mechanism, which rewards guesses within 9 units of the correct assessment. For the Low Uncertainty situations, this "reward window," 18 units across, constituted a very large fraction of the range of possibilities, possibly exerting a pressure to make more "central" guesses in order to (nearly) guarantee *some* payoff; such incentives are clearly weaker in High Uncertainty situations. In retrospect, it might have been better in this regard to elicit assessments with a smaller "reward window."

[18]Of course, in regressions with interaction terms, the overall marginal effect of OBSERVER varies along with the variables with which it is interacted. For each regression, Tables 4 and 6 indicate any values of TIME and any feasible values of RANGEMEAN up to 50 for which the overall marginal effect of OBSERVER is sometimes significant at the $p < 0.05$ level. Overall marginal effects were not

and time-varying, pattern of recipient-observer differences emerges under High Uncertainty (Table 4b). As before, the coefficient of OBSERVER is strongly positive ($p < 0.005$), indicating that observers make higher assessments of proposer offers than recipients do; and, more strongly than before, it is apparent that the recipient-observer gap declines as offers appear likely to have been more generous ($p < 0.005$). The negative coefficient of TIME*OBSERVER ($p < 0.02$) as well as the positive coefficient of TIME*RANGEMEAN*OBSERVER ($p < 0.03$) indicate that this pattern is however mitigated over time, consistent with the original intuitions. A further window on the time evolution of recipient-observer differences is offered by Column (3), which estimates the basic regression specification for the first halves of the experimental sessions only (Periods 1-8), and column (4), which displays results for second-half data (Periods 9-15). Consistent with expectations derived from Column (2), Table 4b demonstrates strong recipient-observer differences in the first halves of experimental sessions, but no statistically significant effect in later periods.[19] A more nuanced period-by-period analysis demonstrates that the measured recipient-observer differences are not due only to anomalous behavior in one or two periods.[20] The observed pattern of temporal decrease in recipient-observer differences is consistent with the original intuition that, despite receiving limited feedback, subjects would nonetheless learn over time about the structure of uncertainty in the experiment.

FIGURE 1 ABOUT HERE

Figure 1 offers a graphical depiction of the overall marginal effect of being an observer, as opposed to a recipient, on assessments of proposers' intentions as the mean of the observed range varies, based on the Table 4b High Uncertainty estimates for the first halves of ultimatum game sessions. The smallest possible value of RANGEMEAN under High Uncertainty is 15 (when $x_{min} = 0$ and $x_{width} = 30$); calculated for values of RANGEMEAN > 50 because very few offers exceeded 50 and therefore little data exists to accurately estimate OBSERVER marginal effects in this regime.

[19]Note that the overall and first-half OBSERVER effect coefficients are substantially *larger* when estimated using the whole dataset rather than following the more conservative strategy of dropping out-of-range assessments that was adopted here. TO EDITORS AND REFEREES: See Table 4bR.

[20]TO EDITORS AND REFEREES: Figure 1R in the Supplemental Appendix shows the point-estimate of the recipient-observer assessments gap for each period estimated separately, fitted when $x_{mean} = 20$. While there is considerable noise, the Figure reflects the intuitions gleaned from Table 4b: there is on average a substantial recipient-observer gap throughout earlier periods, of consistent sign, but a more ambiguous pattern in later periods.

when RANGEMEAN is at this minimum value, observers' assessments of proposers' offers exceed recipients' by about 8 tokens. This gap is substantively quite large, constituting roughly a quarter of the range of possibilities communicated to recipients and observers. As indicated by the negative coefficient for RANGEMEAN*OBSERVER, this gap decreases as RANGEMEAN increases – that is, as the offer appears likelier to have been larger. Taking into account both the direct effect of OBSERVER and the interaction term RANGEMEAN*OBSERVER, the positive overall marginal effect of OBSERVER is significant at the $p < 0.05$ level when RANGEMEAN < 37 and at the $p < 0.10$ level when RANGEMEAN < 40. As depicted in the Figure, the gap declines to statistical insignificance as RANGEMEAN increases beyond this point.

Several additional estimations indicate that this distinctive pattern for High Uncertainty, early period ultimatum games is quite robust in several specific senses. First, strikingly, observers' assessments were higher than recipients' for apparently low offers – but not for offers that appeared likely to be more generous – in all four ultimatum game sessions when these were analyzed separately. Second, this pattern also remains intact under plausible alternative regression specifications. For example, in principle subjects may have inferred that the randomly generated ranges were adjusted to avoid reporting negative values, potentially making subjects' responses qualitatively different when $x_{min} = 0$. However, regression specifications explicitly modeling this possibility return no statistically significant difference in assessments based on whether or not $x_{min} = 0$.[21] Third, as noted above, approximately 5% of the ultimatum game offers were "exotic" or "hyper-fair" offers exceeding 50 (including two outlying offers of exactly 100). Data points corresponding to such atypical, extreme values of an independent variable can sometimes bias estimates of how subjects behave when the independent variable takes on more typical values, given the linearity assumptions implicit in OLS; however, none of a variety of cut rules excluding abnormally high offers from the data make a noticeable difference in the results.

FIGURE 2 ABOUT HERE

The regression results in Tables 4a and 4b, along with Figure 1, provide information about the average effect of strategic role on subjects' assessments of proposers' intentions. Figure 2 provides a

[21]For example, defining a dummy variable XMIN0 that equals 1 when $x_{min} = 0$ but equals 0 when $x_{min} > 0$, a specification of the form $ASSESS_i = \beta_0 + \beta_1 RANGEMEAN_i + \beta_2 OBSERVER_i + \beta_3 RANGEMEAN_i * OBSERVER_i + \beta_4 XMIN0 + \beta_5 XMIN0 * OBSERVER_i + \epsilon_i$, return substantively small and statistically insignificant coefficients β_4 and β_5, and does not induce meaningful changes in the other coefficients reported in the Tables.

finer-grained picture by plotting the cumulative distributions of recipients' and observers' assessments when they were presented with the range $[x_{min}, x_{max}] = [0, 30]$ – the most commonly observed range during ultimatum game play (44 times in 390 periods, or fully 11.3% of the dataset). The cumulative distribution rises more swiftly for recipients than it does for observers, illustrating that recipients exhibit markedly lower assessments of proposers' offers than observers do. However, the Figure also illustrates the high level of heterogeneity in the assessments made both by recipients and by observers. This heterogeneity is quite typical of the data overall.

These results are striking. When uncertainty is sufficiently high, ultimatum game recipients and observers make systematically different assessments of proposers' intentions – even though subjects were *randomly* assigned to these roles and even though the experimental design explicitly *controlled* for the information to which recipients and observers had access. What accounts for the recipient-observer difference? This pattern of behavior calls to mind several plausible interpretations. Because a number of these interpretations are best evaluated in the context of comparing ultimatum *and* dictator game behavior, the paper proceeds to analysis of subjects' post-play assessments in the dictator game before presenting and evaluating interpretations of the ultimatum game role effects.

4.3 Recipients' and Observers' Assessments Do Not Differ in the Dictator Game

Data from the four dictator game sessions was analyzed using the same regression specifications employed in the previous section. The results are contained in Tables 5-6b.

TABLES 5, 6a, 6b AND FIGURE 3 ABOUT HERE

These regression tables do not indicate any meaningful pattern of significant relationships between subjects' strategic roles and their assessments of proposers' intentions. The estimated overall marginal effect of being an observer as opposed to a recipient is consistently both substantively small and statistically insignificant, both for Low and High Uncertainty situations. Examination of period-by-period regression analyses also uncovers no clear pattern; about as many dictator game periods exhibit a negative as a positive recipient-observer gap, there are no obvious time trends in the data, and most of the fitted values are substantively small.[22] Figure 3 offers further intuition about recipients' and observers' assessment behavior by plotting the cumulative distributions of subjects' assessments given the same range $[x_{min}, x_{max}] = [0, 30]$ depicted in ultimatum-game Figure 2 – also a commonly-

[22]TO EDITORS AND REFEREES: See Figure 2R in the Supplemental Appendix, which parallels ultimatum-game Figure 1R. For completeness, Figure 3R parallels ultimatum-game Figure 1.

observed range during dictator game play (44 times in 365 periods, 12.1%). While Figure 2 showed a substantial gap between the distributions of recipients' and observers' assessments, Figure 3 displays a much closer correspondence between recipients' and observers' estimates, reflecting the lack of a statistically-significant difference. However, Figure 3 resembles Figure 2 in one very important regard: the graph indicates a substantial degree of heterogeneity in the assessments made both by recipients and by observers. This high degree of heterogeneity suggests that in the dictator game, like in the ultimatum game, it was far from obvious to recipients and observers how proposers were behaving; the absence of role effects in the dictator game is *not* due merely to all recipients and observers having made identical, very low assessments.

4.4 Why Recipients' and Observers' Assessments Differ in Some Settings But Not in Others

The contrast between the ultimatum and dictator game results is striking. What accounts for it?

Before addressing this question, it is appropriate to consider the extent to which comparisons between ultimatum- and dictator-game assessments are meaningful. After all, it has been argued that factors such as the amount of uncertainty inherent in a given situation will affect the existence and magnitude of strategic role effects on assessment formation, even within a given game. Might the contexts of offers in the ultimatum and dictator games be sufficiently different that comparisons such as this one are inherently misleading?

While such concerns cannot be set aside entirely, the context of assessment formation in the ultimatum and dictator game sessions was strikingly similar in several key respects. The form of proposers' offer decisions, the structure of uncertainty, and the language used to describe the setting to subjects were all identical in both games, as was the method by which recipients' and observers' assessments were elicited. Further, as described in Section 3, the distributions of proposers' offers were extremely similar in the ultimatum and dictator games in terms of mean behavior, modal behavior, and the variance in behavior. The structural commonalities, along with proposers' behavioral similarities, suggest that recipients and observers faced highly similar levels of uncertainty in making their assessments in both games. This point is underscored by the fact that subjects' payoffs associated with the post-play question eliciting assessments were very similar in the two games (averaging 32.8 tokens per period out of a possible 100 for all recipients, and 34.4 for all observers in the ultimatum game; 38.8 for all recipients, and 36.4 for all observers in the dictator game). Notably, the payoff difference across games

for observers, whose task is identical in both games, is statistically insignificant ($p < 0.46$).[23] Further, the overall variance in recipients' and observers' assessments as a function of x_{width} was also remarkably stable across the two games.[24] These observations offer compelling evidence that proposers' behavior was no more predictable for subjects in the dictator game than it was in the ultimatum game, and that recipients and observers faced similar levels of uncertainty in making assessments in both contexts. The widely distributed assessments made by subjects receiving identical information, both in the ultimatum game and in the dictator game (as depicted in Figures 2 and 3), underscore this point. Thus, it is *not* the case that no recipient-observer differences are measured in the dictator game merely because, for example, all subjects in both roles make extremely low assessments all of the time.

Why, then, do recipients make lower assessments of proposers' offers than observers do in the ultimatum game – but not in the dictator game? In the ultimatum game, recipients and observers receive the same information about proposers' offers, but their roles differ in essentially two respects. Recipients have the opportunity to respond to proposers' offers, while observers do not; and recipients' payoffs are affected by proposers' choices, while observers' are not. This observation suggests two psychologically interesting hypotheses about why recipients' and observers' assessments differ in the ultimatum game.[25] First, recipients' assessments may be lower than observers' because of recipients' opportunity to accept or reject offers. This hypothesis is consistent with a psychological mechanism by which the strategic ability to respond to others' behavior may prime individuals to be more negative or critical in evaluating ambiguous evidence about others' intentions than they otherwise would be. In contrast, a second hypothesis suggests that recipients' assessments may be lower because proposers' choices are payoff-relevant for recipients, but not for observers. Such an account is consistent with a psychological model in which recipients' investment in an outcome may affect the way in which they

[23]The ultimatum game payoff for observers is 36.1 if the two deviant subjects described in section 4.1 are dropped.

[24]For comparisons with more than five observations in the Table 4 and Table 6 datasets, the standard deviation in assessments (recipients and observers pooled) was 11.8 in the ultimatum game vs 12.4 in the dictator game for $x_{width} = 40$; 10.1 vs 9.9 for $x_{width} = 35$; 8.8 vs 8.0 for $x_{width} = 30$; 7.0 vs 6.9 for $x_{width} = 25$; and 6.5 vs 5.9 for $x_{width} = 20$.

[25]These actors' assessments could also be thought to differ for substantively uninteresting reasons; for example, they might be an artifact of the assessment elicitation mechanism. Such methodological concerns are discussed shortly.

assess it, perhaps because of an emotional response. A related, but distinct, account might suggest that recipients' affective involvement stimulates conceptions of fairness in a different way than would be the case for more disinterested observers, potentially influencing recipients' expectations about the offers they are likely to receive.[26]

Because recipients and observers differ from one another in *two* key respects in the ultimatum game, it is not possible to make well-supported inferences about the *independent* effects of each of these potential psychological mechanisms on how assessments are made. The dictator game sessions were included in the experimental design in anticipation of such an interpretive confound. Because recipients and observers differ in only *one* key respect in the dictator game, a comparison between ultimatum and dictator game assessments can provide inferential leverage that would be unavailable in analyzing the ultimatum game results on their own.[27] If dictator game assessments were to exhibit a similar pattern of recipient-observer differences to that of ultimatum game assessments, such a result would implicate the payoff-relevance of offers to recipients (but not observers) as the key causal factor. If, on the other hand, dictator game assessments were to exhibit no role effects, this would offer evidence that individuals' assessments can be influenced directly *by the set of strategic options they possess*, controlling for information. The results in Section 4.3, which suggest that the robust recipient-observer differences in the ultimatum game are not present in the dictator game, therefore lend weight to the latter interpretation.

In order to defend and strengthen this inference, it is important to consider potential alternative explanations for the observed patterns of assessments. As a starting point, it is useful to reflect on the nature of the "assessments" that recipients and observers make. These assessments are not, strictly speaking, a direct measure of recipients' and observers' beliefs about what proposers had intended to do. The information these actors receive comes in the form of a range $[x_{min}, x_{max}]$ into which the true

[26]Bolton and Ockenfels (2000) and Fehr and Schmidt (1999) offer compelling theoretical models that implicate fairness concerns in recipient decisions to reject ultimatum-game offers. However, it should be said that this literature offers much clearer guidance in understanding agents' *behavior* than in understanding agents' *expectations* of others' behavior, much less the way in which they would interpret evidence in contexts of uncertainty.

[27]The dictator game was itself originally devised to test different explanations for proposers' typically non-trivial offers to ultimatum-game recipients: altruism *vs* strategic anticipation of recipient rejections (Forsythe et al 1994).

offer, x, fell. Recipients' and observers' beliefs about x would therefore take the form of a probability distribution over $[x_{min}, x_{max}]$. To reliably elicit a meaningful estimate of such a distribution from experimental subjects would be difficult, if not impossible. As such, the approach adopted was to elicit a single integer-valued "assessment" y, and to reward subjects based on the accuracy of these assessments. As such, subjects' assessments reflect, but do not directly measure, their distributions of beliefs about proposers' offers. Different individuals may map beliefs onto assessments differently, based on such factors as their degrees of risk aversion; however, the random assignment of subjects to the recipient and observer roles means that these two groups can be expected to share similar distributions of these factors. The practice of random assignment, along with rewards for accurate assessments, makes these values y reasonable proxies for subjects' beliefs about proposers' intentions, particularly for the purpose of comparing recipient-observer differences across different settings.

While random assignment is, in this way, powerful, it is important to ensure that features of the recipient and observer roles do not themselves greatly affect individual attitudes towards risk aversion. For example, in the ultimatum game, observers receive a fixed game payment of 40 tokens per period, while recipients received, on average, 24.8 tokens per period (standard deviation = 21.7). In the context of the ultimatum game itself, it is imaginable that this difference in game payoffs might induce differential incentives in attempting to secure assessment payoffs. For example, a recipient might think, "If the offer was poor, I can at least ensure a decent payoff by entering a low assessment. If the offer then turned out to be good, then my assessment may have been wrong, but in that case at least my game payoff was good." That is, the ultimatum game results in isolation could potentially be explained, at least in part, by differential monetary incentives in making assessments, based on induced differences in risk aversion or wealth effects across roles. However, comparison of the ultimatum and dictator game results suggests strongly that such an account does not provide a compelling explanation of the ultimatum game results. Observers' game payments were fixed at 40 tokens per round in *both* games; dictator-game recipients received, on average, 23.0 tokens per period (standard deviation = 22.3), practically identical to the ultimatum-game values. Thus, if risk aversion or wealth effects like those described above were the correct account of recipient-observer differences in the ultimatum game, such differences should *also* be observed in the dictator game. The absence of role differences in the dictator game therefore casts considerable doubt on such accounts of the gap in the ultimatum game.[28]

[28]It could also be noted that the temporal decrease in the ultimatum game recipient-observer gap is inconsistent with a risk-aversion account; such an account would, presumably, predict that any role

Another reasonable concern about the measurement of assessments in the ultimatum game is that recipients' assessments may be affected not by the *general fact* of having to respond to offers – the priming argument advanced above – but rather by the *particular choice* to accept or reject that they have *just made.* Psychologically, subjects who reject an offer may in theory subsequently find it more comfortable to believe that the payoff they have foregone is smaller, either to minimize the potential for regret or to downplay the possibility that they may have punished a proposer whose offer had actually been reasonably generous. Similarly, subjects who accept an offer may find it preferable to believe that the payoff they have accepted is larger, either because this prospect is more pleasant to contemplate, or because it may be unpleasant to imagine having been duped into accepting a very poor offer. If such effects exist, then depending on their relative sizes, they might induce systematic differences in how recipients and observers formulate assessments from ambiguous evidence. More clearly, such effects *should* induce a significant correlation between recipients' assessments and the decisions to accept (rather than reject) offers. The data, however, indicate no such correlation. Recall that almost all ultimatum game rejections took place when $x_{min} = 0$; although the range widths varied, different recipients nonetheless had roughly comparable information about proposers' offers in the 151 periods involving $x_{min} = 0$. Yet within this universe of cases, the correlation between recipients' assessments and the decision to accept was -0.028. The absence of any positive relationship between these two variables suggests that the recipient-observer gap in the ultimatum game is not merely an artifact of recipients' reactions to their own specific decisions to accept or reject offers.[29].

It is also worth contemplating the potential role of attentional factors in subjects' assessments of proposers' intentions. Taken on their own, the ultimatum game findings might lend themselves to a rather blunt attentional interpretation. Recipients' thoughts about proposers' offers are relevant both to their game payoffs (because they may affect the decision to accept or reject) as well as to their assessment payoffs; observers, who receive fixed game payoffs, only "need" to think about proposers'

effect would remain static in time. The fact that role differences decrease over time is more consonant with a psychological account involving differential interpretations of evidence under an uncertainty that diminishes over time.

[29]This point is underscored by replicating the regressions in Table 4b, but restricting the analysis only to those recipient-observer dyads for which the recipient *accepted* the proposer's offer. When this is done, the gap between recipients and observers remains intact (and actually grows by a slight and statistically insignificant amount)

offers in order to earn higher assessment payoffs. Might observers simply pay less attention than recipients, leading to a systematic difference in ultimatum-game assessments? Might this explanation also account for the *lack* of a recipient-observer effect in the dictator game, when actors in neither role must make a game decision?

The evidence, however, suggests that the explanatory power of such a blunt attentional model is weak. Factors such as attention are, of course, difficult to assess quantitatively in the context of the experimental data. However, it is notable that observers actually receive slightly *higher* (though not statistically significantly higher) assessment payoffs in the ultimatum game (34.3 tokens on average, versus 32.8 for recipients[30], suggesting that they do not simply disregard the information about proposers' offers because they lack a move in the ultimatum game. This is perhaps unsurprising, given that the $1 potentially at stake for each assessment is, by the standards of the lab experimental literature, a reasonable sum, and given the fact that subjects' assessment payoffs constitute a substantial fraction of their overall payoffs for participation. Some further, indirect, evidence about observers' relative level of attention in the ultimatum game comes in the form of serendipitous data on the speed with which recipients and observers entered their assessments in two of the four ultimatum game sessions.[31] A variety of statistical tests all suggest that there is no significant difference in the amount of time that recipients and observers spent weighing and entering their assessments (for the most straightforward t test, $p < 0.87$).

That these potential alternative explanations fail to gather empirical support adds strength to the inference that recipients' ability to accept or reject offers *causes* the systematic difference between recipients' and observers' assessments of proposers' intentions in the ultimatum game.

[30]The figure for observers grows to 36.1 tokens if the two deviant subjects highlighted in section 4.1 are dropped.

[31]The Z-Tree experimental software (Fischbacher 1999) automatically records the length of time subjects take in recording their choices at each stage of an experimental session, but this function works only when the central experimental server has been rebooted relatively recently. The use of this data on choice timing had not been envisioned prior to carrying out the experimental sessions; serendipitously, the data was recorded for two of the four ultimatum game sessions.

5 Discussion and Conclusion

Do participants in bargaining and conflict form beliefs about one another in systematically different ways than a disinterested observer would, even controlling for the information to which actors have access? More generally, are actors' assessments of others' intentions colored by the strategic situations in which the evaluators find themselves, and if so, why? The laboratory experiment described in this paper suggests that the answer to these questions is yes. By randomly assigning subjects to different roles in experimental games, and controlling for the information to which these subjects had access, it was demonstrated that participants and observers do, under some circumstances, formulate systematically different assessments of another actor's intentions. Comparing patterns of participant and observer assessments *across* different experimental games offered more nuanced evidence as to the potential causes of such strategic role effects on perceptions of others. The evidence supports a striking conclusion: that an actor's assessments of another's intentions can be directly affected by the set of strategic options open to the evaluating actor, even controlling for the actual information to which the evaluator has access. This conclusion is consistent with a psychological mechanism by which actors' evaluations of others may differ based on the mindset induced by looming strategic decisions. In the ultimatum game, for example, the evidence is consistent with an account in which the ability to accept or reject an offer primes decision makers to be more negative or critical in evaluating ambiguous evidence about a counterpart actor's intentions.

While this experiment hints at new directions, the details of the hypothesized psychological mechanism by which the power to accept or reject can "prime" individuals to be more critical or negative in assessing proposers' offers can only be uncovered through further research.[32] Just as analysis of results from the *dictator* game made possible causal inferences about the psychological mechanisms underlying participant-observer differences in the *ultimatum* game, future studies examining the way

[32]For example, the ability to accept or reject may affect the salience of different kinds of interpretations of ambiguous evidence in actors' minds; it may put actors in a different emotional state, affecting the way in which they form judgments; or the power to respond may in itself directly affect individuals' senses of what they deserve in a given setting, thereby influencing judgments by changing the way in which actors think about fairness. For many applications in international relations, it may well be enough to know that the "priming" effect exists without nailing down the psychological micro-mechanism; for other applications, it may be necessary to know more.

in which actors in different roles form inferences about others, in different strategic settings and with different kinds of information, will both provoke and answer new questions offering fresh insights into a foundational question in the literatures on bargaining and conflict: how individuals actually assess the motivations of their counterparts in practice.

One potentially fruitful direction for theoretical inquiry concerns the implications of the recipient-observer gap for proposers' incentives. In the dictator game, recipients and observers share a common strategic position, and according to the experimental results, they also share common assessments of proposers' intentions given the same information. Neither has any say in the turn of events; their perceptions of a proposer's intentions are, in a strategic sense, beside the point from the proposer's point of view. In the ultimatum game, however, recipients and observers are strategically distinct, and their assessments of proposers' intentions differ accordingly, even given the same information. Here recipients' assessments could, potentially, matter to proposers in a strategic sense, because an actor's perceptions of a counterpart's intentions may influence his or her willingness to accept a bargain or to extend trust. Interestingly, a recipient in the ultimatum game could potentially *benefit* from making systematically lower assessments than he or she would as a neutral observer; in some circumstances, a proposer who anticipates such a recipient-observer gap might be induced to offer *more* because of it, because a lower assessment of an offer may make rejection more likely in many settings. A psychological mechanism by which the power to accept or reject primes actors to assess more negatively may therefore serve as a de facto *credible commitment* device. Such a device would be all the more credible for functioning at the level of *perception*, taking the matter out of the hands of conscious reasoning by the actor it serves.[33] In this connection, it is notable that the recipient-observer gap in the ultimatum game is significant *only* for offers that appear to have been relatively poor – exactly the offers for which the possibility of rejection comes into question as a practical matter.

Of course, extrapolating the results of any laboratory experiment to real-world settings of political interest is an uncertain endeavor. Nonetheless, if actors' strategic roles can have a substantial effect on the assessments they make even in such a stylized environment, in which the information offered is ambiguous but clearly specified, in which subjects are *randomly* assigned to experimental roles, and in which levels of emotional investment are relatively low, it would seem natural to hypothesize that actors' roles likely *would* have an effect as well in more politically realistic settings, which are more richly contextual, in which the structure of uncertainty is likely to be more complex, and in which actors are

[33]Citation to a published theoretical paper on this point by the author is suppressed.

more likely to be emotionally invested in outcomes. In spite of this, of course, external validity concerns will always be present. But a willingness to take such concerns in stride opens the door to a high degree of *internal* validity, based on the ability to randomly assign individual subjects to different roles while carefully controlling for the information to which they have access. These features of the experimental design allow for clean causal inferences about the effects of strategic role on belief formation that could not reasonably be demanded of observational data, case studies, or surveys involving social groups with ongoing memberships. As always, a broad and deep understanding of political processes is best pursued using multiple methodological approaches; laboratory experimentation, survey research, and the qualitative study of empirical cases can, in tandem, through and in spite of their diverse strengths and weaknesses, teach us more than any one approach could on its own.

The experimental results and the psychological interpretation advocated here have important implications for our understanding of bargaining and conflict. First and foremost, if participants in conflict are indeed, under some knowable set of circumstances, systematically more pessimistic about one another's intentions than a neutral observer would be, this finding would have direct implications for theories of conflict onset, conflict duration, and the potential roles for impartial mediators in peace processes. But it also has implications for the design of optimal *institutions* for conflict resolution. If actors' assessments of one another are directly affected by the mere fact of having, or lacking, particular strategic options, then this suggests that some institutional procedures for conflict resolution may be more successful than others are in efficiently translating good intentions into *perceptions* of good intentions – and ultimately, into good relations. Further research – experimental, empirical, and theoretical – into the psychology of judgment formation in conflict and bargaining may ultimately aid not only in our understanding of conflicts, but in finding their solutions.

6 References

Bolton, Gary E. and Axel Ockenfels. 2000. "ERC - A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90(1): 166-193.

Camerer, Colin. 2003. *Behavioral Game Theory.* Princeton: Princeton University Press.

Chatman, Celina M. and William von Hippel. 2001. "Attributional Mediation of In-Group Bias." *Journal of Experimental Social Psychology* 37: 267-272.

Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation."

Quarterly Journal of Economics 114: 817-868.

Fischbacher, Urs. 1999. "z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual." Working Paper Nr. 21, Institute for Empirical Research in Economics, University of Zurich.

Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior* 6: 347-369.

Gerber, Alan and Donald P. Green. 1999. "Misperceptions about Perceptual Bias." *Annual Review of Political Science* 2: 189-210.

Guth, Werner, R Schmittberger and B Schwarz. 1982. "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization* 3: 367-88.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations.* London: Lawrence Erlbaum Associates.

Jervis, Robert. 1976. *Perception and Misperception in International Politics.* Princeton: Princeton University Press.

Jervis, Robert, Richard Ned Lebow, and Janice Gross Stein, eds. 1985. *Psychology and Deterrence.* Baltimore: Johns Hopkins University Press.

Jones, Edward E. 1979. "The Rocky Road from Acts to Dispositions." *American Psychologist* 34(2): 107-117.

Jones, Edward E. and Keith E. Davis. 1965. "From Acts to Dispositions: the Attribution Process in Person Perception," in *Advances in Experimental Social Psychology* 2, L. Berkowitz, ed., pp. 219-266. New York: Academic Press.

Jones, Edward E. and Richard E. Nesbitt. 1971. *The Actor and the Observer: Divergent Perceptions of the Causes of Behavior.* Morristown NJ: General Learning Press.

Kelley, Harold H. 1967. "Attribution Theory in Social Psychology." in David Levine (ed.), Nebraska Symposium on Motivation 15: 192-240.

Kulik, James. 1983. "Confirmatory Attribution and the Perpetuation of Social Beliefs." *Journal of Personality and Social Psychology* 44(6): 1171-1181.

McDermott, Rose. 2004. *Political Psychology in International Relations.* Ann Arbor: University of Michigan Press.

Mercer, Jonathan. 1996. *Reputation and International Politics.* Ithaca: Cornell University Press.

Mitzkewitz, Michael and Rosemarie Nagel. 1993. "Experimental Results on Ultimatum Games with Incomplete Information." *International Journal of Game Theory* 22(2): 171-98.

Rapoport, Amnon and James A. Sundali. 1996. "Ultimatums in Two-Person Bargaining with One-Sided Uncertainty: Offer Games." *International Journal of Game Theory* 25: 475-494.

Rapoport, Amnon, James A. Sundali, and Darryl A. Seale. 1996. "Ultimatums in Two-Person Bargaining with One-Sided Uncertainty: Demand Games." *Journal of Economic Behavior and Organization* 30: 173-196.

Rosenberg, Shawn W. and Gary Wolfsfeld. 1977. "International Conflict and the Problem of Attribution." *Journal of Conflict Resolution* 21(1): 75-103.

Ross, Lee. 1977. "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process," in *Advances in Experimental Social Psychology* 10, L. Berkowitz, ed., pp. 173-220. New York: Academic Press.

Taber, Charles S. and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3): 755-769.

Taylor, Shelley and Judith Hall Koivumaki. 1976. "The Perception of Self and Others." *Journal of Personality and Social Psychology* 33(4): 408.

Tetlock, Philip E. 1985. "Accountability: A Social Check on the Fundamental Attribution Error." *Social Psychology Quarterly* 48(3): 227-236.

Tillman, Wayne and Charles Carver. 1980. "Actors' and Observers' Attributions for Success and Failure." *Journal of Experimental Social Psychology* 16(1): 18-32.

Van der Plight, Joop and J. Richard Eiser. 1983. "Actors and Observers' Attributions, Self-Serving Bias and Positivity Bias." *European Journal of Social Psychology* 13: 95-104.

Table 1. Experimental Sessions

Session	Number of Subjects	Number of Periods
Ultimatum-1	21	15
Ultimatum-2	21	15
Ultimatum-3	21	15
Ultimatum-4	15	15
Dictator-1	15	10
Dictator-2	21	15
Dictator-3	21	15
Dictator-4	21	15

Table 2. Actual Offers in the Ultimatum and Dictator Games

.	Ult. Game	Ult. Game	Dict. Game	Dict. Game
x	Number of Obs.	Cumulative Distn	Number of Obs.	Cumulative Distn
0-5	109	28.0%	115	31.5%
6-10	31	35.9%	39	42.2%
11-15	11	38.7%	27	49.6%
16-20	11	41.5%	16	54.0%
21-25	30	49.2%	29	61.9%
26-30	31	57.2%	26	69.0%
31-35	28	64.4%	18	74.0%
36-40	43	75.4%	13	77.5%
41-45	30	83.1%	13	81.1%
46-50	47	95.1%	48	94.3%
51+	19	100%	21	100%
TOTAL	**390**	**-**	**365**	**-**

Table 3. Recipients' and Observers' Assessments of Proposers' Offers in the Ultimatum Game: Separate Estimations for Different Uncertainty Conditions

Dependent Variable: Assessment of Proposer's Offer

OLS Regression with Standard Errors Clustered on Individual Subjects

Out-of-Range Assessments Dropped

.	$x_{width} = 15$	$x_{width} = 20$	$x_{width} = 25$	$x_{width} = 30$	$x_{width} = 35$	$x_{width} = 40$
.	(N=26)	(N=141)	(N=203)	(N=183)	(N=140)	(N=55)
.	20 clusters	50 clusters	49 clusters	48 clusters	49 clusters	35 clusters
RANGEMEAN	.	0.925(0.030)***	0.833(0.035)***	0.834(0.045)***	0.813(0.068)***	0.584(0.103)***
OBSERVER	-1.006(1.627)	0.291(3.002)	-2.111(2.673)	6.228(3.013)**	8.653(4.691)*	7.200(10.865)
RANGEMEAN*OBSERVER	.	0.001(0.066)	0.116(0.070)	-0.080(0.062)	-0.185(0.095)*	-0.099(0.183)
Constant	7.733(0.875)***	2.366(1.361)*	4.457(1.643)***	2.680(2.271)	4.051(3.458)	13.167(6.060)**
R^2	0.018	0.886	0.8445	0.7554	0.5583	0.4118

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

Tables 4a and 4b. Recipients' and Observers' Assessments of Proposers' Offers: the Ultimatum Game

Dependent Variable: Assessment of Proposer's Offer

OLS Regressions with Standard Errors Clustered on Individual Subjects

Out-of-Range Assessments Dropped

Table 4a. Low Uncertainty

	(1) All Periods (N=370) 50 clusters	(2) All Periods (N=370) 50 clusters	(3) Periods 1-8 (N=203) 49 clusters	(4) Periods 9-15 (N=167) 50 clusters
RANGEMEAN	0.882(0.026)***	0.882(0.026)***	0.871(0.025)***	0.898(0.043)***
OBSERVER	-0.915(2.372)	-3.045(2.781)	-3.349(2.718)	2.217(2.657)
RANGEMEAN*OBSERVER	0.068(0.058)	0.134(0.075)*	0.129(0.063)**	-0.036(0.073)
TIME*OBSERVER	.	0.320(0.200)	.	.
TIME*RANGEMEAN*OBSERVER	.	-0.011(0.007)	.	.
Constant	2.965(1.209)**	2.965(1.213)**	3.491(1.409)**	2.388(1.541)
R^2	0.8749	0.8758	0.8996	0.8219
OBS. overall marg. effect (**)?	Never	Never	+ ($48 \leq R \leq 50$)	Never

Table 4b. High Uncertainty

	(1) All Periods (N=378) 50 clusters	(2) All Periods (N=378) 50 clusters	(3) Periods 1-8 (N=196) 50 clusters	(4) Periods 9-15 (N=182) 50 clusters
RANGEMEAN	0.795(0.040)***	0.795(0.040)***	0.829(0.052)***	0.754(0.052)***
OBSERVER	6.698(3.193)**	12.314(3.945)***	11.515(3.592)***	1.935(3.527)
RANGEMEAN*OBSERVER	-0.105(0.057)*	-0.225(0.075)***	-0.222(0.072)***	0.017(0.072)
TIME*OBSERVER	.	-0.732(0.290)**	.	.
TIME*RANGEMEAN*OBSERVER	.	0.016(0.007)**	.	.
Constant	4.140(2.277)*	4.140(2.283)*	2.891(2.579)	5.572(2.553)**
R^2	0.6656	0.6697	0.6679	0.6707
OBS. overall marg. effect (**)?	+ ($R \leq 26$)	+ ($R \leq 37$, $T \leq 9$)	+ ($R \leq 36$)	Never

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

Table 5. Recipients' and Observers' Assessments of Proposers' Offers in the Dictator Game: Separate Estimations for Different Uncertainty Conditions

Dependent Variable: Assessment of Proposer's Offer

OLS Regression with Standard Errors Clustered on Individual Subjects

Out-of-Range Assessments Dropped

.	$x_{width} = 15$	$x_{width} = 20$	$x_{width} = 25$	$x_{width} = 30$	$x_{width} = 35$	$x_{width} = 40$
.	(N=4)	(N=93)	(N=161)	(N=188)	(N=190)	(N=89)
.	4 clusters	45 clusters	47 clusters	50 clusters	52 clusters	44 clusters
RANGEMEAN	.	1.010(0.067)***	0.997(0.045)***	0.988(0.054)***	1.117(0.060)***	0.956(0.126)***
OBSERVER	-2.000(1.581)	-0.634(2.686)	-0.450(2.875)	-1.486(2.986)	3.531(3.210)	-5.190(6.629)
RANGEMEAN*OBSERVER	.	-0.027(0.090)	0.008(0.060)	-0.019(0.079)	-0.117(0.085)	0.076(0.182)
Constant	8.500(0.500)***	-2.847(1.945)	-3.715(2.084)*	-3.124(2.334)	-9.204(2.206)***	-3.507(4.960)
R^2	0.4444	0.8536	0.8943	0.8397	0.8002	0.6184

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

Tables 6a and 6b. Recipients' and Observers' Assessments of Proposers' Offers: the Dictator Game

Dependent Variable: Assessment of Proposer's Offer

OLS Regression with Standard Errors Clustered on Individual Subjects

Out-of-Range Assessments Dropped

Table 6a. Low Uncertainty

	(1) All Periods (N=258) 50 clusters	(2) All Periods (N=258) 50 clusters	(3) Periods 1-8 (N=139) 50 clusters	(4) Periods 9-15 (N=167) 50 clusters
RANGEMEAN	0.988(0.036)***	0.988(0.036)***	0.951(0.051)***	1.020(0.044)***
OBSERVER	-0.786(2.068)	1.232(2.394)	-1.139(2.151)	-0.516(2.730)
RANGEMEAN*OBSERVER	0.008(0.047)	-0.062(0.061)	0.025(0.065)	-0.008(0.057)
TIME*OBSERVER	.	-0.270(0.216)	.	.
TIME*RANGEMEAN*OBSERVER	.	0.010(0.005)*	.	.
Constant	-3.016(1.570)*	-3.016(1.576)*	-1.997(1.606)	-3.884(2.102)*
R^2	0.8915	0.8923	0.8744	0.9066
OBS. overall marg. effect (**)?	Never	Never	Never	Never

Table 6b. High Uncertainty

	(1) All Periods (N=467) 52 clusters	(2) All Periods (N=467) 52 clusters	(3) Periods 1-8 (N=273) 52 clusters	(4) Periods 9-15 (N=194) 52 clusters
RANGEMEAN	1.041(0.055)***	1.041(0.055)***	0.972(0.051)***	1.137(0.073)***
OBSERVER	0.159(2.914)	4.080(3.452)	-0.002(3.189)	0.434(3.344)
RANGEMEAN*OBSERVER	-0.051(0.073)	-0.126(0.088)	-0.012(0.075)	-0.109(0.094)
TIME*OBSERVER	.	-0.559(0.257)**	.	.
TIME*RANGEMEAN*OBSERVER	.	0.010(0.008)	.	.
Constant	-5.911(2.304)**	-5.911(2.309)**	-4.185(2.447)*	-8.122(2.629)***
R^2	0.7810	0.7830	0.7581	0.8142
OBS. overall marg. effect (**)?	Never	Never	Never	Never

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

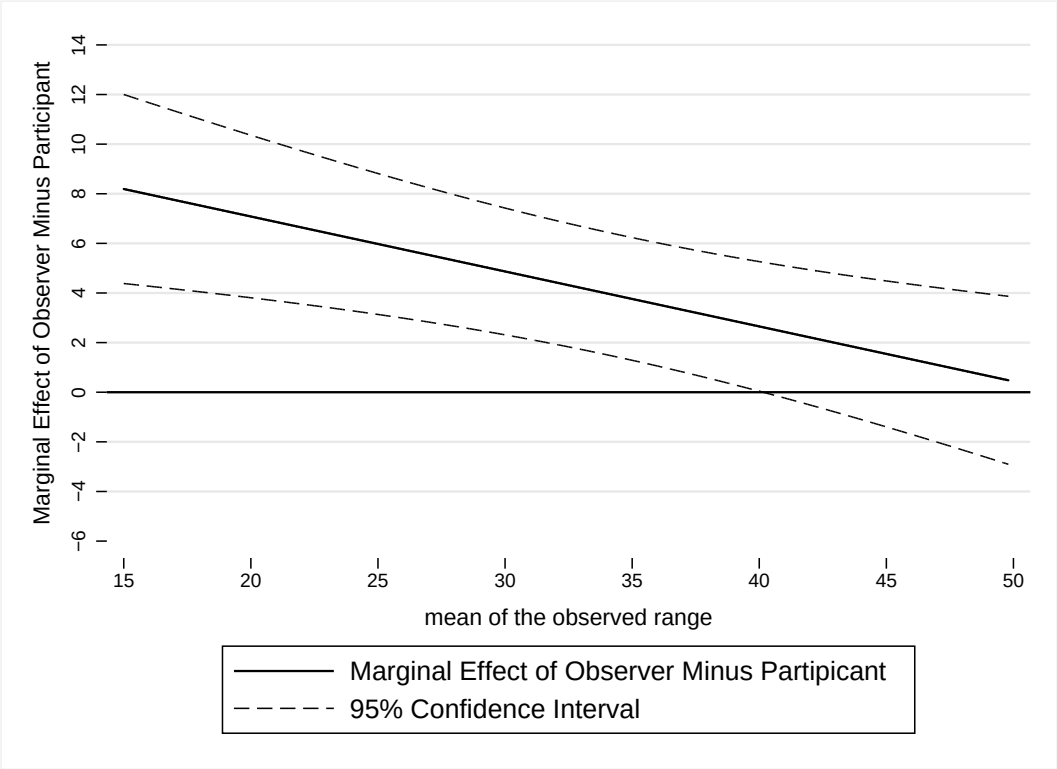

Figure 1: The Overall Marginal Effect of Being an Observer (Rather than a Recipient) on Assessments of Proposers' Offers Under High Uncertainty During the First Half of Ultimatum Game Sessions.

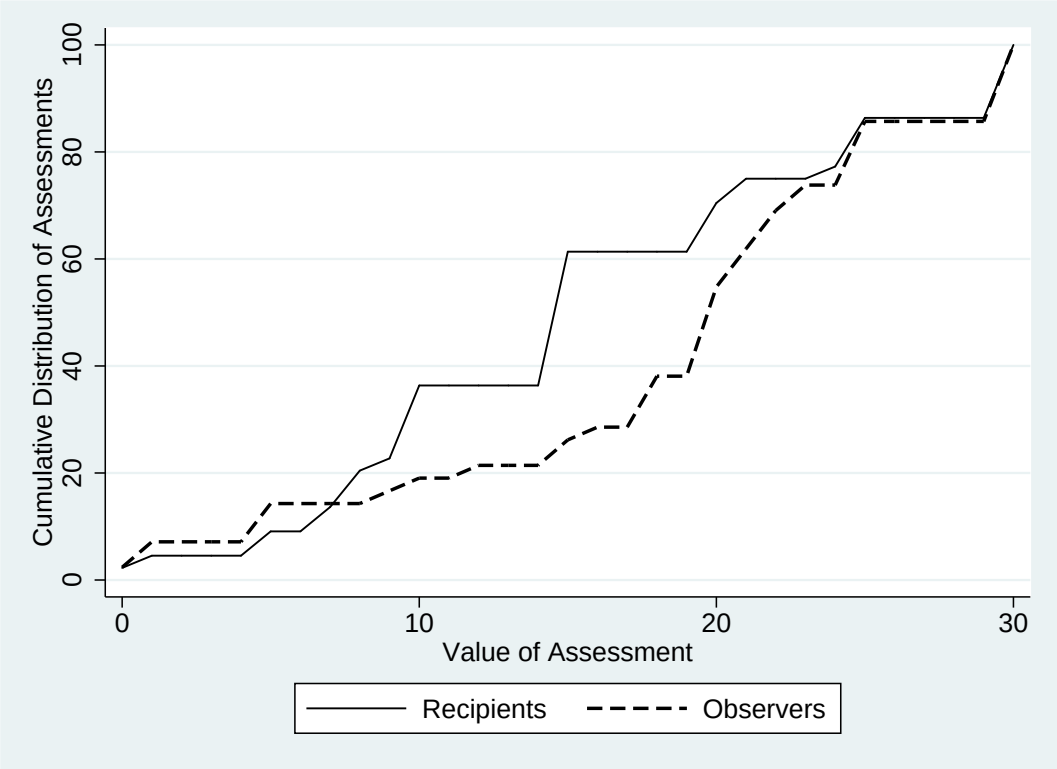

Figure 2: Cumulative Distribution of Recipients' and Observers' Assessments of Proposers' Offers During Ultimatum Game Sessions, when $x_{min} = 0$ and $x_{max} = 30$.

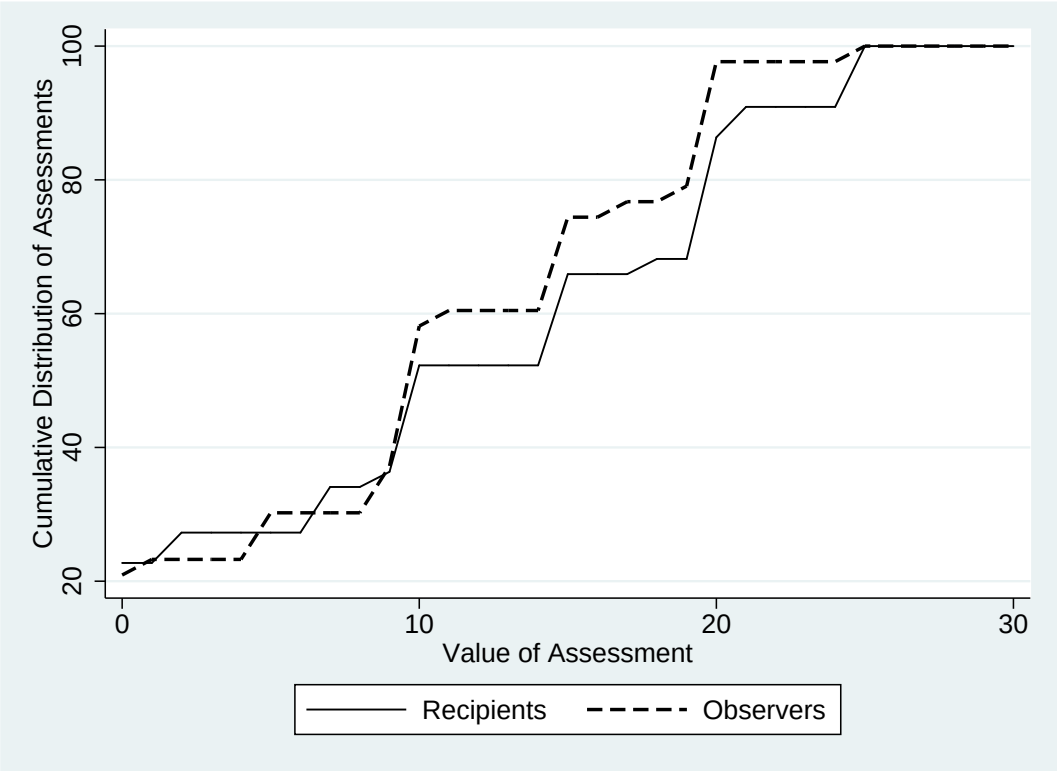

Figure 3: Cumulative Distribution of Recipients' and Observers' Assessments of Proposers' Offers During Dictator Game Sessions, when $x_{min} = 0$ and $x_{max} = 30$ (to be compared with Figure 2).

Referees' Supplemental Appendix

for

Do Participants and Observers Assess Intentions Differently During Bargaining and Conflict? An Experiment

Table 3R. Recipients' and Observers' Assessments of Proposers' Offers in the Ultimatum Game: Separate Estimations for Different Uncertainty Conditions

Dependent Variable: Assessment of Proposer's Offer

OLS Regression with Standard Errors Clustered on Individual Subjects

Full Dataset

	$x_{width} = 15$ (N=30) 22 clusters	$x_{width} = 20$ (N=144) 52 clusters	$x_{width} = 25$ (N=208) 51 clusters	$x_{width} = 30$ (N=194) 50 clusters	$x_{width} = 35$ (N=146) 51 clusters	$x_{width} = 40$ (N=58) 37 clusters
RANGEMEAN	.	0.925(0.030)***	0.833(0.035)***	0.834(0.045)***	0.813(0.068)***	0.584(0.103)***
OBSERVER	10.4(6.899)	0.882(3.250)	-0.294(3.287)	9.503(4.082)**	18.580(7.593)**	7.098(9.912)
RANGEMEAN*OBSERVER	.	0.010(0.081)	0.078(0.082)	-0.131(0.081)	-0.406(0.166)**	-0.078(0.162)
Constant	7.733(0.871)***	2.366(1.360)*	4.457(1.642)***	2.680(2.269)	4.051(3.455)	13.167(6.046)**
R^2	0.0820	0.8301	0.7843	0.6090	0.3453	0.4159

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

Tables 4aR and 4bR. Recipients' and Observers' Assessments of Proposers' Offers: the Ultimatum Game

Dependent Variable: Assessment of Proposer's Offer

OLS Regressions with Standard Errors Clustered on Individual Subjects

Full Dataset

Table 4aR. Low Uncertainty

	All Periods (N=382) 52 clusters	All Periods (N=382) 52 clusters	Periods 1-8 (N=206) 51 clusters	Periods 9-15 (N=176) 52 clusters
RANGEMEAN	0.882(0.026)***	0.882(0.026)***	0.871(0.025)***	0.898(0.043)***
OBSERVER	2.224 (3.153)	0.962(5.417)	-0.130(3.692)	5.349(3.788)
RANGEMEAN*OBSERVER	0.007(0.073)	0.061(0.123)	0.076(0.075)	-0.113(0.103)
TIME*OBSERVER	.	0.197(0.607)	.	.
TIME*RANGEMEAN*OBSERVER	.	-0.009(0.016)	.	.
Constant	2.965(1.209)**	2.965(1.212)**	3.491(1.408)**	2.388(1.539)
R^2	0.7607	0.7613	0.7898	0.6963

Table 4bR. High Uncertainty

	All Periods (N=398) 52 clusters	All Periods (N=398) 52 clusters	Periods 1-8 (N=210) 52 clusters	Periods 9-15 (N=188) 52 clusters
RANGEMEAN	0.795(0.040)***	0.795(0.040)***	0.829(0.052)***	0.754(0.052)***
OBSERVER	11.687(4.428)**	23.312(7.556)***	19.133(5.614)***	3.800(3.999)
RANGEMEAN*OBSERVER	-0.198(0.082)**	-0.414(0.128)***	-0.353(0.101)***	-0.025(0.085)
TIME*OBSERVER	.	-1.575(0.642)**	.	.
TIME*RANGEMEAN*OBSERVER	.	0.029(0.011)**	.	.
Constant	4.140(2.276)*	4.140(2.281)*	2.891(2.576)	5.572(2.551)**
R^2	0.5079	0.5268	0.4695	0.5886

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

Table 5R. Recipients' and Observers' Assessments of Proposers' Offers in the Dictator Game: Separate Estimations for Different Uncertainty Conditions

Dependent Variable: Assessment of Proposer's Offer

OLS Regression with Standard Errors Clustered on Individual Subjects

Full Dataset

	$x_{width} = 15$ (N=4) 4 clusters	$x_{width} = 20$ (N=94) 45 clusters	$x_{width} = 25$ (N=162) 47 clusters	$x_{width} = 30$ (N=190) 50 clusters	$x_{width} = 35$ (N=190) 52 clusters	$x_{width} = 40$ (N=90) 45 clusters
RANGEMEAN	.	0.996(0.070)***	0.962(0.064)***	0.988(0.054)***	1.117(0.060)***	0.913(0.132)***
OBSERVER	-2.000(1.581)	-1.201(2.761)	-2.323(3.801)	-0.675(3.008)	3.531(3.210)	-5.817(6.727)
RANGEMEAN*OBSERVER	.	-0.013(0.092)	0.0434(0.075)	-0.019(0.082)	-0.117(0.085)	0.118(0.186)
Constant	8.500(0.500)***	-2.279(2.048)	-1.843(3.245)	-3.124(2.333)	-9.204(2.206)***	-2.880(5.092)
R^2	0.4444	0.8452	0.8406	0.8075	0.8002	0.5954

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

Tables 6aR and 6bR. Recipients' and Observers' Assessments of Proposers' Offers: the Dictator Game

Dependent Variable: Assessment of Proposer's Offer

OLS Regression with Standard Errors Clustered on Individual Subjects

Full Dataset

Table 6aR. Low Uncertainty

	All Periods (N=260) 50 clusters	All Periods (N=260) 50 clusters	Periods 1-8 (N=140) 50 clusters	Periods 9-15 (N=120) 45 clusters
RANGEMEAN	0.964(0.045)***	0.964(0.045)***	0.909(0.066)***	1.011(0.044)***
OBSERVER	-2.016(2.515)	0.001(2.793)	-3.173(3.111)	-0.991(2.751)
RANGEMEAN*OBSERVER	0.032(0.054)	-0.039(0.067)	0.067(0.078)	0.002(0.058)
TIME*OBSERVER	.	-0.270(0.216)	.	.
TIME*RANGEMEAN*OBSERVER	.	0.010(0.005)*	.	.
Constant	-1.786(2.125)	-1.786(2.133)	0.037(2.762)	-3.409(2.129)
R^2	0.8516	0.8523	0.7951	0.9028

Table 6bR. High Uncertainty

	All Periods (N=470) 52 clusters	All Periods (N=470) 52 clusters	Periods 1-8 (N=276) 52 clusters	Periods 9-15 (N=194) 52 clusters
RANGEMEAN	1.035(0.056)***	1.035(0.056)***	0.961(0.053)***	1.137(0.073)***
OBSERVER	0.419(2.928)	5.108(3.494)	0.459(3.206)	0.434(3.344)
RANGEMEAN*OBSERVER	-0.120(0.093)	-0.126(0.088)	-0.002(0.078)	-0.109(0.094)
TIME*OBSERVER	.	-0.676(0.260)**	.	.
TIME*RANGEMEAN*OBSERVER	.	0.010(0.008)	.	.
Constant	-5.853(2.318)**	-5.853(2.323)**	-4.091(2.478)	-8.122(2.629)**
R^2	0.7649	0.7683	0.7305	0.8142

(***) denotes $p < 0.01$; (**) denotes $p < 0.05$; (*) denotes $p < 0.10$.

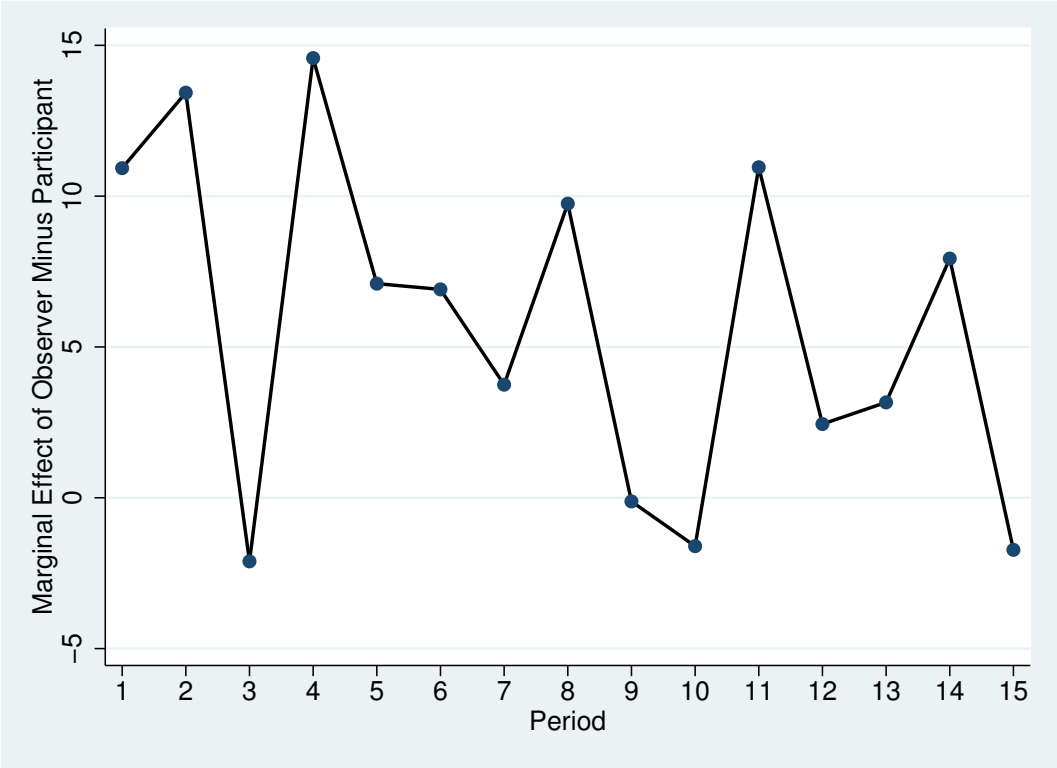

Figure 1: Figure 1R. The Overall Marginal Effect of Being an Observer (Rather than a Recipient) on Assessments of Proposers' Offers During Ultimatum Game Sessions Under High Uncertainty, Period-by-Period, Estimated Value When $x_{mean} = 20$.

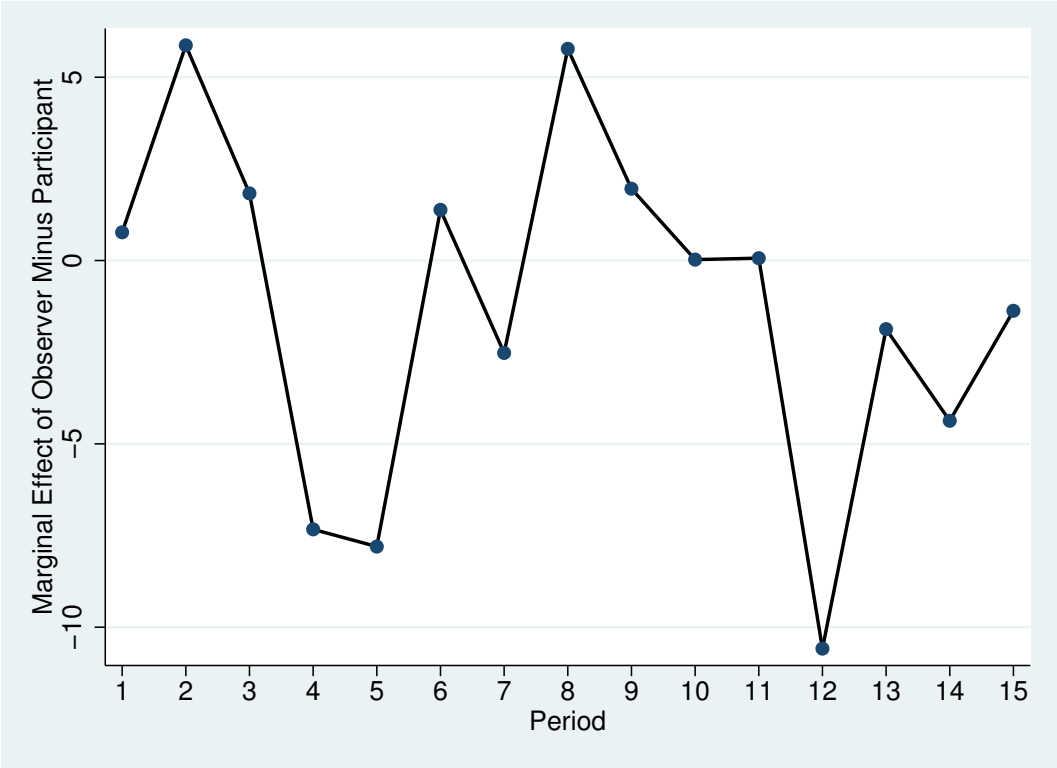

Figure 2: Figure 2R. The Overall Marginal Effect of Being an Observer (Rather than a Recipient) on Assessments of Proposers' Offers During Dictator Game Sessions Under High Uncertainty, Period-by-Period, Estimated Value When $x_{mean} = 20$.

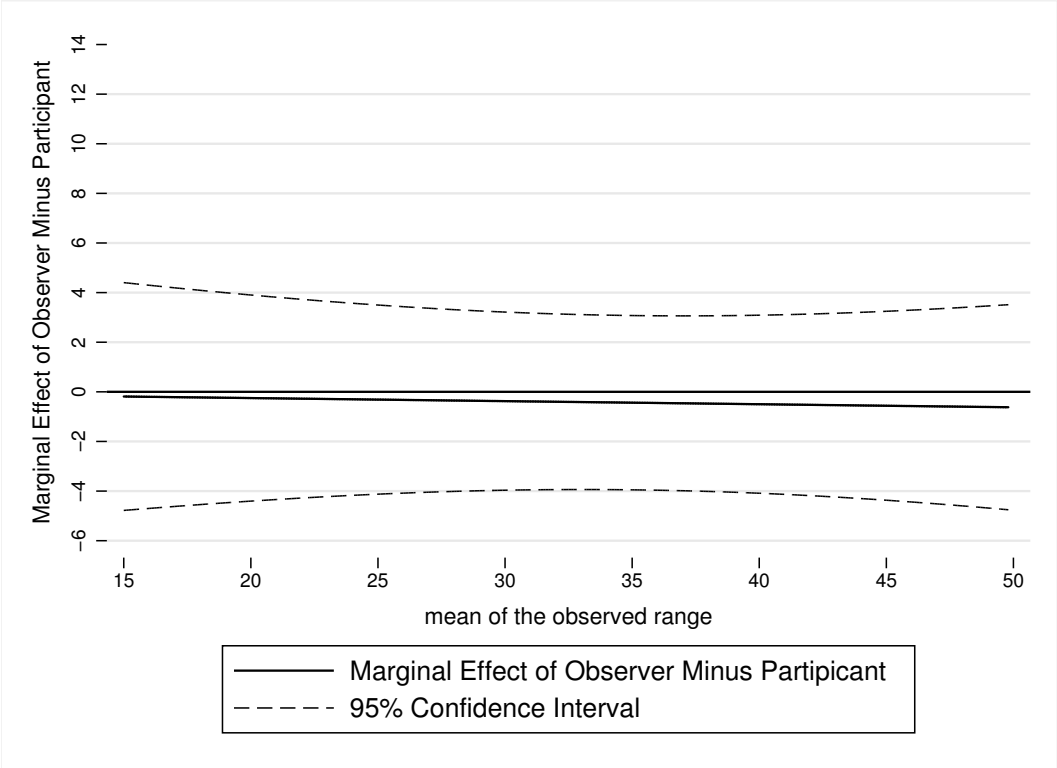

Figure 3: Figure 3R. The Overall Marginal Effect of Being an Observer (Rather than a Recipient) on Assessments of Proposers' Offers Under High Uncertainty During the First Half of Dictator Game Sessions (to be compared with Figure 1).

Reviewers' Appendix: Initial Instructions to Subjects in all Roles

[Underlined contents in brackets were specific to the ultimatum game sessions.]
[**Bold contents in brackets were specific to the dictator game sessions.]**
All other contents were common to both ultimatum and dictator game sessions.

Instructions

Introduction

This is an experiment on decision making. In the following experiment you will make a series of choices. At the end of the experiment, you will be paid depending on the specific choices that you made during the experiment and the choices made by other people. If you follow the instructions and make appropriate decisions, you may make an appreciable amount of money. Please listen carefully to the instructions, and take a moment to turn off any cell phones or pagers that you may have with you.

During the course of the experiment, you will have the opportunity to earn "tokens" that will be converted into dollars at the end of the experiment. The conversion rate is:

100 tokens = 1 dollar

In a moment, the proctor will launch the experimental software, and the computer will randomly assign you to one of three roles: Role 1, Role 2, or Role 3. The role to which you are assigned will remain fixed throughout the procedure described in these instructions. When your role assignment appears on your screen, please take a moment to write down your assigned role in the blank below, using the red pen that has been provided to you. Please do not click "OK" until you are instructed to do so.

MY ROLE IS: _____

You will participate in a number of independent *periods*. At the beginning of each period, you will be assigned, at random, into a group of three people among whom interactions will take place. Each group will contain one person in Role 1, one person in Role 2, and one person in Role 3. At the end of each period, groups will be dissolved, and at the beginning of the next period, you will again be assigned at random into a new group, consisting of one person in each of the three Roles. Remember that your own assigned Role remains fixed throughout this process.

All of your interactions with others will be through the computer terminals at which you are sitting, and your true identity will never be revealed to any other person in the laboratory.

All of the independent *periods* have the same general structure. In each period, there are two separate *stages*.

- In the first stage of each period, [the people in Role 1 and Role 2 both make choices] **[the person in Role 1 makes a choice]** (the nature of which will be described below). [These choices affect] **[This choice affects]** the earnings received by the people in Role 1 and Role 2, but [they do not affect] **[it does not affect]** the earnings of the person in Role 3.
- In the second stage of each period, the people in Role 1, Role 2, and Role 3 will all be asked to make a guess about a choice made by one of their counterparts in their group. Each person will then receive earnings based on his or her own answer to the question that he or she was asked.

Details of what happens in the two stages are discussed below.

First Stage

In the first stage, the person in Role 3 receives fixed earnings of **40** tokens that are unaffected by the choices made by the people in the other Roles. The earnings of the people in Role 1 and Role 2 are determined by a process in which [both the person in Role 1 and the person in Role 2 make] **[the person in Role 1 makes]** a choice. The specific sequence of events is as follows.

- First, the person in Role 1 must propose a division of 100 tokens between him- or herself and the person in Role 2. Specifically, the person in Role 1 offers to give ***x*** tokens to the person in Role 2 while keeping the rest of the tokens, **100-*x***, for him- or herself. ***x*** must be a whole number of tokens (i.e., no fractions). For example, if the person in Role 1 offers to give ***x*** = 40 tokens to the person in Role 2, this entails keeping **100-*x*** = 60 tokens for him- or herself. As another example, if the person in Role 1 offers to give ***x*** = 70 tokens to the person in Role 2, this entails keeping **100-*x*** = 30 tokens for him- or herself.
- The value of ***x*** chosen by the person in Role 1 is not directly observed by the other people in the group. Instead, they are told only that ***x*** falls within a particular "range" that is observed by everyone in the group. For example, if the person in Role 1 offers to give ***x*** = 40 tokens to the person in Role 2, everyone could be informed that ***x*** is between 20 and 40; or they could be informed that ***x*** is between 30 and 70; or they could be informed that ***x*** is between 40 and 65. That is, the true value of ***x*** may be equal to the lower number in the range, or it may be equal to the upper number in the range, or it may be any whole number that lies in between the lower and upper numbers in the range.
- **[Finally, after everyone in the group has been informed of the "range" in which *x* falls, the payoffs of the person in Role 1 and the person in Role 2 are determined in the following way. The person in Role 2 receives *x* tokens, while the person in Role 1 receives 100-*x* tokens.]**

- [Finally, after everyone in the group has been informed of the "range" in which ***x*** falls, the person in Role 2 must then decide whether to accept or to reject the offer made by the person in Role 1. If the person in Role 2 accepts, then he or she receives ***x*** tokens, while the person in Role 1 receives **100-*x*** tokens. If the person in Role 2 rejects, then he or she receives **0** tokens, and the person in Role 1 also receives **0** tokens.]

Remember that the earnings of the person in Role 3 for the first stage are simply fixed at **40** tokens.

Second Stage

In the second stage, the people in Role 1, Role 2, and Role 3 must all answer an on-screen question giving their guess about a choice made by one of their counterparts in their group. Regardless of your specific Role, you will then receive earnings based on the degree to which your guess is accurate, in a way that will be described on your screen. You will be asked the same question, and your earnings will be calculated in the same way, in the second stage of each of the independent periods.

Conclusion

The same process will be repeated in all of the independent periods, each of which will consist of the two stages described above. Remember that you will be randomly rematched into a new group of three people at the beginning of each period. Your earnings from each stage of each period will be added to your total earnings. The computer will keep track of your earnings as they accumulate, but you will not receive immediate feedback as to what your earnings are.

Please remain silent until the end of the last period. If you have any questions, please ask them at this time.

Reviewers' Appendix: On-Screen Instructions to Recipients and Observers Regarding Post-Play Questions

Make your best guess as to the value of *x* chosen by the person in Role 1. It is in your interests for your guess to be as close as possible to the actual value of *x*.

Specifically, if your guess falls between 0 and 9 units away from the actual value of x, you will receive a payoff of 100 tokens MINUS 10 tokens for every "unit of distance" between your guess and the actual value of x.

If your guess falls 10 or more units away from the actual value of x, you will receive no payoff for your guess.

Your guess as to the value of x chosen by the person in Role 1:_____