THE ROLE OF BOUNDED RATIONALITY AND IMPERFECT INFORMATION IN SUBGAME PERFECT IMPLEMENTATION—AN EMPIRICAL INVESTIGATION

Philippe Aghion
Harvard University

Ernst Fehr
University of Zurich

Richard Holden
University of New South Wales

Tom Wilkening
University of Melbourne

Abstract

In this paper we conduct a laboratory experiment to test the extent to which Moore and Repullo's subgame perfect implementation mechanism induces truth-telling, both in a setting with perfect information and in a setting where buyers and sellers face a small amount of uncertainty regarding the good's value. We find that Moore–Repullo mechanisms fail to implement truth-telling in a substantial number of cases even under perfect information about the valuation of the good. Our data further suggests that a substantial proportion of these lies are made by subjects who hold pessimistic beliefs about the rationality of their trading partners. Although the mechanism should—in theory—provide incentives for truth-telling, many buyers in fact believe that they can increase their expected monetary payoff by lying. The deviations from truth-telling become significantly more frequent and more persistent when agents face small amounts of uncertainty regarding the good's value. Our results thus suggest that both beliefs about irrational play and small amounts of uncertainty about valuations may constitute important reasons for the absence of Moore–Repullo mechanisms in practice. (JEL: D23, D71, D86, C92)

1. Introduction

Subgame perfect implementation has attracted much attention since it was introduced by Moore and Repullo (1988). A main reason for this success is the remarkable

The editor in charge of this paper was Juuso Välimäki.

Acknowledgments: We owe special thanks to Michael Powell and Eric Maskin. We also thank Christopher Engel, Oliver Hart, Martin Hellwig, Andy Postlewaite, Klaus Schmidt, Larry Samuelson, and seminar participants at the 2010 Asian-Pacific ESA Conference (Melbourne, Australia), Bocconi, Chicago Booth, Harvard, MIT, Stanford, the IIES in Stockholm, the Max Planck Institute in Bonn, UNSW and University of Queensland for helpful comments. We gratefully acknowledge the financial support of the Australian Research Council including ARC Future Fellowship FT130101159 (Holden) and ARC Discovery Early Career Research Award DE140101014 (Wilkening), the University of Melbourne Faculty of Business and Economics, and the European Research Council grant on the Foundations of Economic Preferences (Fehr).

E-mail: paghion@fas.harvard.edu (Aghion); ernst.fehr@econ.uzh.ch (Fehr); richard.holden@unsw.edu.au (Holden); tom.wilkening@unimelb.edu.au (Wilkening)

Journal of the European Economic Association 2018 16(1):232–274 DOI: 10.1093/jeea/jvx026

property that almost any social choice function can be implemented as the *unique* subgame perfect equilibrium of a suitably designed dynamic mechanism.[1] This was perceived as a substantial improvement over Nash implementation, which suffered from two main limitations: first, it would allow only a certain class of social choice rules to be implemented, those which are "Maskin Monotonic" (Maskin 1977; Maskin 1999); roughly speaking, Nash implementation does not permit the implementation of social choice rules that involve distributional concerns between the agents. Second, Nash implementation typically involves multiple equilibria, so that even if a desirable equilibrium exists, an undesirable one may too.

A common objection to subgame perfect implementation mechanisms, however, is that they are hardly observed in practice. This in turn raises the question as to why one does not observe them. A first type of answer, discussed by Selten (1975), van Damme (1984), and Fudenberg, Kreps, and Levine (1988), is that the behavioral assumptions embedded in subgame perfection may not be a good approximation of actual behavior. Another type of answer[2] is that subgame perfect implementation is not robust to arbitrarily small deviations from common knowledge.

In this paper we use a laboratory experiment to test the extent to which the Moore–Repullo mechanism implements truth-telling in practice, both in a setting with perfect information and in a setting where buyers and sellers do not share common knowledge about the good's valuations. We implement three treatments: one with perfect information about the value of the good (we refer to it as the no-noise treatment); one with 5% imperfect information (i.e., traders receive information about the good's valuation that is 95% correct); and one with 10% imperfect information (traders have information that is 90% correct). We also conducted a robustness check with only 1% imperfect information to examine whether even very small deviations from complete information can cause serious failures in inducing truth-telling.

Our environment is taken from Hart and Moore (2003) where a seller is about to receive a buyer-specific good of either high or low quality. Before learning the value of the good, the buyer and seller would like to write a contract where the buyer pays a high price if the good is of high quality and a low price if the good is of low quality. However, the quality of the good is not verifiable by a third-party court and thus a state-dependent contract cannot be directly enforced.

Although the state is not verifiable, public announcements can be recorded and used in legal proceedings. Thus the two parties can in principle write a contract that specifies trade prices as a function of announcements made by the buyer. If the buyer always tells the truth, then his announcement can be used to set state-dependent prices. One way of doing this is to implement a mechanism that allows announcements to be challenged by the seller and to punish the buyer any time he is challenged. If the

[1] Subgame perfect implementation also assumes that individuals are sequentially rational and that transfers of any size are allowed.

[2] See Aghion et al. (2012), henceforth AFHKT.

seller challenges only when the buyer has told a lie, then the threat of punishment will ensure truth-telling.

The key challenge of developing the implementation mechanism is to construct a set of rules such that the seller has an incentive to challenge lies but to prevent the seller from challenging the buyer when he has in fact told the truth. The subgame perfect implementation (SPI) mechanism we consider accomplishes this by having a seller's challenge trigger two actions: a punishment, in the form of a fine, and a counter-offer. This counter-offer is structured so that if the buyer was lying he will accept the counter-offer and if he was telling the truth he will reject it. By conditioning additional award and punishments to the seller based on whether the counter-offer was accepted or rejected, the mechanism can prevent sellers from abusing their power by challenging when the buyer had indeed told the truth.

Thus, overall the mechanism has three stages: the announcement stage at which the buyer announces the value of the good, the challenge stage at which the seller has the option to challenge the buyer's announcement, and a counter-offer stage at which the buyer can accept or reject the counter-offer in case the seller has made a challenge. For the SPI mechanism to induce truth-telling at the announcement stage, the later stages must be structured so that (i) buyers have an incentive to accept counter-offers after a lie and to reject counter-offers after the truth and (ii) sellers have an incentive to challenge lies and not challenge truthful announcements. When experimenting with the SPI mechanism outlined above under full information, we find that the mechanism is very successful in inducing these behaviors. In line with what the theory would predict, buyers always reject counter-offers after a truthful announcement and accept counter-offers over 90% of the time after a lie; sellers challenge lies over 90% of time and challenge truthful announcements in less than 5% of cases.

Surprisingly, however, the mechanism in our full information treatment fails to induce truth-telling in a substantial number of cases. Despite correct pecuniary incentives, buyers who observe a high quality good lie over 30% of the time and about 10% of buyers lie in every period. Based on beliefs data, these lies appear to be due to buyers who are pessimistic about the rationality of the sellers and fear that truthful announcements will be challenged.

To better understand the extent to which beliefs are playing a role, we ran an additional treatment where we elicited incentive compatible beliefs using an elicitation mechanism similar to the BDM mechanism (Becker, DeGroot, and Marschak 1964), which was first developed in Savage (1971).[3] We find that not only do the majority of individuals who lie believe that they have a higher expected pecuniary payoff for lying than for telling the truth, but the majority of individuals who tell the truth also hold these beliefs. This finding is due primarily to a large majority of buyers who believe that truth-telling may be challenged. Thus paradoxically, although the mechanism is designed to induce truth-telling based on pecuniary incentives, the mechanism is in

[3] Variations of this elicitation method have been used by DuCharme and Donnell (1973), Grether (1981), Allen (1987), and Holt (2006). It is shown by Karni (2009) that the mechanism induces truthful reporting of beliefs for rational agents with any von Neumann–Morgenstern utility function.

fact associated with beliefs that render lying profitable for the buyers—even for the majority of buyers who tell the truth. Thus, it appears that a substantial amount of the observed truth-telling is not due to the mechanism but to the buyers' intrinsic preferences for honesty.

Note that our results indicate that the mechanism does not simply fail to induce truth-telling because the subjects generally fail to understand backward induction. In fact, subjects' behavior at the challenge stage and at the counter-offer stage is very close to the backward induction prediction. Rather, the mechanism fails because it generates a specific fear in buyers that they will be challenged in case of truth-telling.

Next, we analyze how the SPI mechanism performs in the presence of imperfect information. More specifically, we introduce two noise treatments where we give buyers and sellers imperfect signals about the underlying quality of the good which are correct either 90% or 95% of the time.[4] We find that the introduction of noise increases the proportion of buyers who announce a low value with a high signal by 15–25 percentage points relative to the no-noise treatment. These buyer lies are persistent in the noise treatment and do not diminish with experience. Further, the introduction of noise causes a significant change in buyers' beliefs; they are now much more likely to believe that lying will not be challenged. Finally, we find that the introduction of noise also exacerbates a pattern that we already observed in the perfect information treatment: the buyers have even more pessimistic beliefs about being challenged after truth-telling.

In a further experiment, we study how the introduction of even small amounts of noise impacts the mechanism. In a treatment where individuals are given the correct signal 99% of the time, we find that the introduction of noise increases buyer lies to the levels observed in the 95% noise treatment. Thus, even very small deviations from common knowledge can have a big effect on the outcome of the mechanism.

The buyers' beliefs that even truthful announcements will be challenged by the sellers seems to play an important role for the mechanism's failure to induce truth-telling both under complete and incomplete information. But does this belief indeed cause buyers' lies? To examine this question, we also study what Moore (1992) refers to as a simple mechanism where we prevent buyers from being challenged if they announce a high valuation for the good. This simple mechanism can implement the first best in our setting but would not function in more complicated environments where both parties must announce truthfully. In a treatment of this mechanism with no noise, the new mechanism dramatically reduces the proportion of buyer lies, providing direct evidence that strategic uncertainty is driving most of the lies in the no-noise treatments. With noise, however, buyer lies continue to be common. Overall, our findings suggest

[4] When noise is introduced, theory predicts that the truth-telling equilibrium is eliminated and both a mixed strategy equilibrium and two pure strategy equilibrium emerge. As discussed in Section 3.3.1, we concentrate on the mixed strategy equilibrium in the paper because the pure strategy equilibrium are based on out-of-equilibrium beliefs that are unlikely to exist in our experimental environment. See Appendix A for a broader discussion of equilibria that may exist in the noise treatments.

that small amounts of private information do indeed lead to large deviations from truth-telling and significantly more lies than under perfect information.

This paper relates to several strands of literature. It first contributes to the literature on mechanism design and more specifically on subgame perfect implementation (Moore and Repullo 1988; Maskin 1999; Maskin and Tirole 1999b; Chung and Ely 2003) by pointing to two main sources for the failure of SPI mechanisms: namely, common knowledge about the rationality of the other players, and (small) deviations from common knowledge in payoffs. In particular we show that beliefs about the irrationality of the trading partner undermine the SPI mechanism even in the case of perfect information about the good's value. This in turn suggests that future work should concentrate on the design and examination of mechanisms that are robust to deviations from perfect information and perfect rationality.[5] Our results also point to a preference for truth-telling that causes some individuals to go against their belief-based pecuniary payoffs and make truthful announcements. This result suggests that it may be possible to design more efficient implementation mechanisms that utilize these preferences for honesty.[6]

Second, our paper contributes to the debate on the foundations of incomplete contracts. In their influential 1986 paper, Grossman and Hart argued that in contracting situations where states of nature are observable but not verifiable, asset ownership (or vertical integration) can help limit ex post hold-up and thereby encourage ex-ante investments (see Grossman and Hart 1986). However, in subsequent work, Maskin and Tirole (1999a,b) used subgame perfect implementation to show that the nonverifiability of states of nature can be overcome using a 3-stage subgame perfect implementation mechanism which induces truth-telling by all parties as the unique equilibrium outcome. Our paper sheds light on why such mechanisms are not observed in practice, which in turn helps to explain why vertical integration or the allocation of control rights matter.

Third, our paper also contributes to the experimental literature on implementation. Sefton and Yavas (1996) study extensive-form Abreu–Matsushima mechanisms that vary in the number of stages used and find that incentive-compatible mechanisms with 8 and 12 stages perform worse than a mechanism with 4 stages that is not incentive compatible. Katok, Sefton, and Yavas (2002) study both simultaneous and sequential

[5] Irrationality and a lack of common knowledge about the rationality of others appear to be important forces in a number of other experimental settings. For example, in the formation of asset bubbles, Lei, Noussair, and Plott (2011) find that bubbles are driven in part by irrationality whereas Cheung, Hedegaard, and Palan (2014) find evidence that common knowledge of rationality is important. In environments with two-sided asymmetric information where no trade is predicted, Carrillo and Palfrey (2011) find that naïve belief formation may help to explain violation of the no-trade theorem whereas Angrisani et al. (2011) find that violations decrease over time in environments where feedback is informative. The systematic under estimation of the rationality of others is similar to results in Huck and Weizsäcker (2002) who find that beliefs about the play of others are distorted toward the uniform prior.

[6] Our result that many individuals tell the truth when their monetary gain from truth-telling is negative is related to the literature on lying aversion (Gneezy 2002; Sanchez-Pages and Vorsatz 2007; Ederer and Fehr 2009).

versions of the Abreu–Matsushima mechanism and conclude that individuals use only a limited number of iterations of dominance and steps of backward induction. Our paper studies mechanisms that require only two steps of backward induction based on the findings in these papers.

An extensive experimental literature also exists looking at efficiency of implementation mechanisms in the public goods provision problems,[7] Solomon's dilemma problems, and problems involving the selection of arbitrators. Ponti, Gantner, López-Pintado, and Mongtgomery (2003) study two implementation mechanisms for the Solomon's dilemma problem and motivate their paper by a concern that first movers in a two-stage mechanism may take suboptimal actions if they fear that second movers will behave irrationally. Our paper finds evidence in support of this channel. Discussing the search for good mechanisms for the selection of arbitrators, de Clippel, Eliaz, and Knight (2014) argue that one desiderata in the search for good mechanisms is that a "mechanism has as few stages as possible so that backward induction is relatively 'simple' to execute". Our paper provides empirical support for this criterion by showing that eliminating unneeded branches of a mechanism can lead to large improvements in its performance.

Most closely related to our paper is Fehr, Powell, and Wilkening (2014) who show that reciprocity considerations may cause the SPI mechanism to fail. By contrast, in this paper we intentionally designed our mechanism and environment so that reciprocity is unlikely to play a role in the no-noise treatment,[8] and concentrate instead on how incomplete information and other forces such as irrational beliefs and strategic uncertainty may affect SPI.

The remaining part of the paper is organized as follows. Section 2 presents the simple model that guides our experimental design. Section 3 describes the experiment and hypotheses. Section 4 presents the experimental results under perfect and imperfect information. Section 5 concludes by suggesting broader implications from our experiment and avenues for future research.

2. Theoretical Motivation

In this section we present a simple example that will guide our experimental design.

[7] Chen and Plott (1996), Chen and Tang (1998), and Healy (2006) study learning dynamics in public good provision mechanisms. Andreoni and Varian (1999), Falkinger et al. (2000), and Chen and Gazzale (2004) study two-stage compensation mechanisms that build on work from Moore and Repullo (1988), whereas Harstad and Marrese (1981, 1982), Attiyeh, Franciosi, and Isaac (2000), Arifovic and Ledyard (2004), and Bracht, Figuières, and Ratto (2008) study the voluntary contribution game, Groves–Ledyard, and Falkinger mechanisms respectively. Masuda, Okano, and Saijo (2014) study approval mechanisms and emphasize the need for implementation mechanisms to be robust to multiple reasoning processes and behavioral assumptions.

[8] We deliberately chose parameters in our experiment that made reciprocal behavior very costly and thus very unlikely to occur in the no-noise treatment.

2.1. Common Knowledge

The following example is based on Hart and Moore (2003).[9] There are two parties, a B(uyer) and a S(eller) of a single unit of an indivisible good. If trade occurs then B's payoff is $V_B = \theta - p$, where θ is the value of the good and p is the price. Ss payoff is just $V_S = p$.

The good can be of either high (the state is $\theta = \theta^H$) or low quality ($\theta = \theta^L$). If it is high quality then B values it at 70, and if it is low quality then B values it at 20. Before θ is realized both parties would prefer to trade at a price $p(\theta) = \theta/2$. This price always ensures that trade occurs when it is efficient and splits the surplus evenly between the buyer and the seller in all states of the world so that inequity aversion does not influence the desire for trade.

The value θ is *observable* and common knowledge to both parties but *nonverifiable* by a court. The assumption that the value θ is nonverifiable implies that no contract can be written that is credibly contingent on θ. However, truthful revelation of θ can be achieved through the following Moore–Repullo (MR) mechanism that can indirectly generate the desired price schedule (see also the extensive form representation of the mechanism in Figure 1):

(1) B announces either "high" or "low". If "high" and S does not "challenge" B's announcement, then B pays S a price equal to 35 and the game then ends.
(2) If B announces "low" and S does not "challenge" B's announcement, then B pays a price equal to 10 and the game ends.
(3) If S challenges B's announcement then:
   (a) B pays a fine of $F = 25$ to T (a third party).
   (b) B is made a counter-offer for the good at a price of 75 if his announcement was "high" and a price of 25 if his announcement was "low".
   (c) If B accepts the counter-offer then S receives the fine $F = 25$ from T (and also the counter-offer price from B) and the game ends.
   (d) If B rejects the counter-offer then S pays $F = 25$ to T. S also gives the good to T who destroys it and the game ends.

When the true value of the good is common knowledge between B and S, individuals are rational, and there is common strong belief in rationality, this mechanism yields truth-telling as the unique subgame-perfect equilibrium. The logic of this equilibrium is that the initial-prices, counter-offer prices, and fines are constructed so that if B and S are commonly known to be sequentially rational, B only has an incentive to announce "high" if $\theta = \theta^H$ and "low" if $\theta = \theta^L$. As can be seen in the extensive form representation of the mechanism shown in Figure 1, for this to be true, the mechanism must satisfy three conditions.

[9] This original example is also reported in Aghion and Holden (2011).

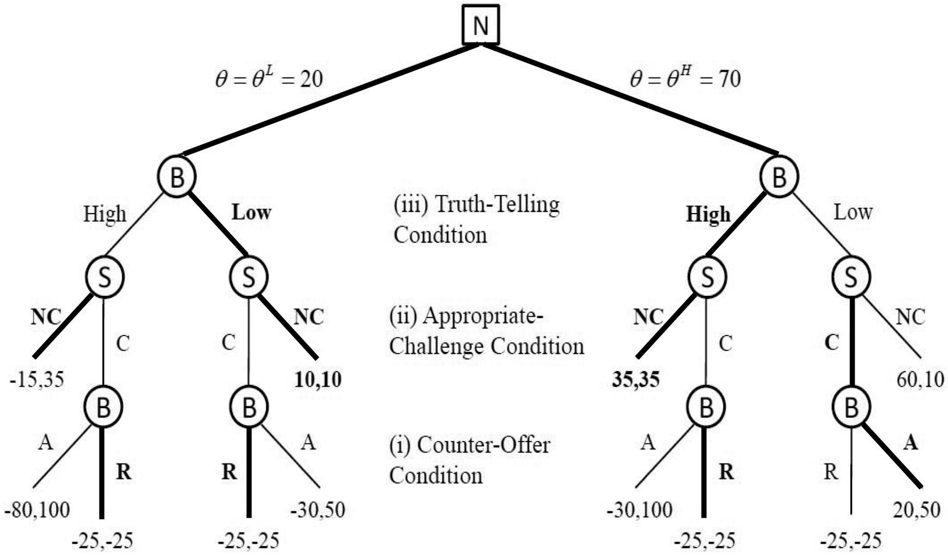


FIGURE 1. Extensive form representation of the Moore–Repullo mechanism. First, Nature determines the value of the good after which the buyer announces whether the value is high or low. The seller observes the value of the good and the buyer's announcement and can then challenge (C) or not challenge (NC) the announcement. In the final stage, the buyer can accept (A) or reject (R) the counteroffer. The unique subgame perfect equilibrium (when θ is common knowledge) is illustrated by bold branches.

(i) *Counter-Offer Condition. B* must prefer to accept any counter-offer for which he has announced "low" when $\theta = \theta^H$. *B* must prefer to reject any counter-offer for which he has announced "low" when $\theta = \theta^L$ or for which he announced "high".

(ii) *Appropriate-Challenge Condition. S* must prefer to challenge an announcement of "low" when $\theta = \theta^H$ and must prefer not to challenge an announcement of "low" when $\theta = \theta^L$. *S* must prefer to never challenge "high".

(iii) *Truth-Telling Condition. B* must prefer to announce "low" if $\theta = \theta^L$ and "high" if $\theta = \theta^H$.

We discuss these conditions in greater detail in Section 3.3.

2.2. The Failure of Truth-Telling Under (Small) Informational Perturbations

We now introduce a small common p-belief perturbation from common knowledge about the valuation θ. We assume (i) the players have a common prior μ, (ii) $\mu(\theta = \theta^H = 70) = 0.5$, and (iii) $\mu(\theta = \theta^L = 20) = 0.5$.[10] Each player receives an

[10] AFHTK consider a more general setting with an arbitrary prior. However, to map closest to the experiment, we develop the theoretical part with the same values, priors, and error distributions as those used in the actual experiment in the next section.

independent draw from a signal structure with two possible signals: s^H or s^L, where s^H is a high signal where θ equals 70 with probability $1 - \varepsilon$, and s^L is a low signal where θ is equal to 20 with probability $1 - \varepsilon$. We use the notation s^H_B (resp. s^L_B) to indicate that B received the high signal s^H (resp. the low signal s^L).

With a small common p-belief perturbation there is no equilibrium in pure strategies in which the buyer and seller always report truthfully. To see this, suppose instead that such an equilibrium exists, and further suppose that B gets signal s^L_B, announces "low", and is challenged. Under a truth-telling equilibrium, the buyer's belief is that his signal and the seller's signal are incorrect with equal probability, and thus the expected value of the good is 45. As this is above the counter-offer price of 25, the buyer has an incentive to purchase regardless of his signal.

Anticipating the acceptance of challenges with a low signal and "low" announcement, the seller now has an incentive to challenge even if his signal is s^L_S. It follows that there does not exist an equilibrium where all parties are truth-telling in pure strategies. For slight changes in the environment, a similar pattern can hold in the case of a buyer who receives signal s^H_B and is considering whether to make the "high" or "low" announcement. In this case, under the truth-telling equilibrium, the seller will be unsure as to the value of the good and may not challenge the announcement if she believes the buyer will reject the counter-offer.

3. The Experiment

3.1. The Subgame-Perfect Implementation Game

At the center of our experimental design is a computerized version of the subgame perfect implementation game we discussed in the previous section. In each of twenty periods, a buyer is matched with a seller and randomly assigned one of two sealed containers.[11] One container is worth 70 Experimental Currency Units (ECU) to the buyer whereas the other container is worth 20 ECU.[12] Containers are selected with equal probability and both the buyer and seller do not initially know which container has been chosen while trading.

Each of the two containers is filled with red and blue balls whose composition changes by treatment:

(1) *No-Noise Treatment.* In the no-noise treatment, the container worth 70 ECU is filled with 20 red balls and 0 blue balls. The container worth 20 ECU is filled with 20 blue balls and 0 red balls.

[11] Subjects are randomly assigned the role of a buyer or of a seller and remain in this role throughout the experiment.

[12] The exchange rate of ECU to Australian dollars was at a rate of 10 ECU = 1 AUD.

(2) *5% Noise Treatment.* In the 5% noise treatment, the container worth 70 ECU is filled with 19 red balls and 1 blue ball. The container worth 20 ECU is filled with 19 blue balls and 1 red ball.

(3) *10% Noise Treatment.* In the 10% noise treatment, the container worth 70 ECU is filled with 18 red balls and 2 blue balls. The container worth 20 ECU is filled with 18 blue balls and 2 red balls.

At the beginning of each period, one of the balls in the container assigned to the buyer is randomly drawn and secretly shown to the seller. This ball is put back into the container and a second ball is randomly drawn for the buyer. These signals provide perfect information regarding the container being traded in the no-noise treatment and almost perfect information in the 5% and 10% noise treatments.[13]

Unlike the seller's draw, we do not immediately show the buyer his draw from the container. Before the buyer knows the color of his ball he is asked to make an announcement concerning the value of the container for the case in which the ball drawn for him is red or blue. He may announce a value of either 70 ECU or 20 ECU in each of the two cases. This strategy method gives us a complete set of announcement data in each period that precludes changes in the frequency of lies over time due to random assignment of signals to different subsets of buyers. The strategy method also allows for a complete panel of choices that improves our ability to control for heterogeneity across individuals.[14]

After making choices for both possible signals, the color of the ball drawn is revealed to the buyer and his declared strategy for this ball color is implemented by the computer. This results in a single buyer announcement that is used in the following stages.

The announcement of the buyer (but not his complete strategy) is next seen by the seller as well as a computerized arbitrator who acts as the implementation mechanism. After observing the announcement, the seller has the option of accepting the announcement or calling the arbitrator. If the seller accepts the announcement, trade occurs at a price equal to $1/2$ of the announcement. If, however, the seller elects to call the arbitrator, the buyer is immediately charged a fine of 25 ECU and the game continues on to the arbitration response stage.

In the arbitration response stage, the buyer is given a counter-offer by the computerized arbitrator that is based on his initial announcement. If he announced a value of 70 ECU, the arbitrator gives a counter-offer of 75 ECU. If he announced a value of 20 ECU, the arbitrator gives a counter-offer of 25 ECU.

[13] In the control quiz, subjects are asked to calculate the likelihood of the other party having the same color ball as them in each treatment. For the no-noise treatment we announce in the verbal summary that "if you see a red ball, you know with 100% certainty that your matched partner has also seen a red ball. Likewise, if you see a blue ball, you know with 100% certainty that your matched partner has also seen a blue ball." For the noise treatments we announce the probability that both parties observe the same signal.

[14] We ran four pilot sessions without the strategy method. The lying rates in these pilot sessions were slightly higher than those reported in the results section, but the treatment effect is similar.

If the buyer accepts the counter-offer, trade occurs at the counter-offer price. In this case the seller is given the 25 ECU that was previously charged as a fine to the buyer.[15] If, however, the buyer rejects the counter-offer, no trade occurs and the seller is also charged a fine of 25 ECU yielding a loss of 25 ECU for both parties. Note that the structure of fines ensures that under full information the subgame-perfect equilibrium is unique.

In the event that trade occurs, the actual value of the container is revealed and the profits of the buyer and seller are realized based on the value of the container, the price, and any fines. The profits of each individual are calculated after each period.

In addition to action profiles of the implementation mechanism, we also elicited beliefs about the likelihood of actions of the other party. Likelihoods were recorded using a 4-point Likert scale (Never/Unlikely/Likely/Always). The belief elicitation was done in each period directly after the buyer or seller took their action. For a buyer, we elicited the likelihood that the seller would challenge an announcement of 20 ECU and 70 ECU in each period given his observed signal, and we did so right after the buyer made her announcement decision but before discovering the seller's action. For a seller, we elicited the likelihood that the buyer would reject the seller's challenge given the observed announcement and the seller's signal.

We did not pay subjects for their beliefs because in the main sessions we were primarily interested in the behavioral data. If we had compensated subjects for both their beliefs and their actions, risk averse subject could have found it optimal to hedge risk by stating beliefs that differ from their true estimates—a possibility that is discussed in more detail in Blanco et al. (2010). Moreover, we ran four additional sessions where we did not elicit beliefs to check whether belief elicitation affects behavior. In these sessions subjects faced 5% noise for 10 periods and then no noise for 10 periods. We find no behavioral differences between these control sessions and the main sessions with the same treatment ordering and with belief elicitation. In particular, the distribution of buyer announcements after a high signal neither differs in the 5% noise treatments (Mann–Whitney–Wilcoxen test, p-value $= 0.53$) nor in the no-noise treatments (Mann–Whitney–Wilcoxen test, p-value $= 0.34$).

3.2. *Experimental Design and Protocols*

As shown in Table 1, our experimental design utilizes a within-subjects design in which each subject is exposed to 10 periods of the no-noise treatment and 10 periods of one of the two noise treatments. Our main experiments consisted of 16 sessions: eight with a 5% noise level and eight with a 10% noise level. We conducted half the sessions starting with the no-noise treatment and switching to the noise treatment in period 11. We reversed the order of the two treatments in the remaining sessions. Each session

[15] As discussed in the next section, sellers were given the entire fine to maximize the incentive of sellers to make challenges across a large range of potential beliefs.

TABLE 1. Treatments and observations—10 periods per treatment.

	Treatment 1	Treatment 2	Number of subjects
Session 1–4	No Noise	5% Noise	88
Session 5–8	5% Noise	No Noise	84
Session 9–12	No Noise	10% Noise	90
Session 13–16	10% Noise	No Noise	86

contained between 20 and 24 subjects who were evenly divided between buyers and sellers at the beginning of the experiment. Buyers and sellers were matched with each other at most once in each of the two treatments.

All of the experiments were run in the Experimental Economics Laboratory at the University of Melbourne in September and October of 2009. The experiments were conducted using the programming language z-Tree (Fischbacher 2007). All of the 348 participants were undergraduate students at the University, who were randomly invited from a pool of more than 3000 volunteers using ORSEE (Greiner 2004). An additional 340 participants were recruited in follow-up sessions conducted in 2010 and 2013.

Upon arrival to the laboratory, participants were divided into buyers and sellers and asked to read the instructions. To be as fair as possible to the mechanism, the instructions described the game in detail, explaining each possible signal, announcement, and arbitration action profiles in order to make the payoff consequences of a challenge and the rejection/acceptance of a challenge transparent. The instructions also included a summary table that showed the payoff consequences of each combination of container value, announcements, challenges, and responses to challenges for both the buyer and the seller. The instructions then ended with a set of practice questions that tested subjects' understanding of the signal valuations and the payoff consequences of accepting or rejecting counter-offers after a lie and after a truthful announcement. Once the answers of all participants were checked, the experimenter read aloud a summary of the instructions. The purpose of the summary was to ensure that the main features of the experiment were common knowledge amongst the participants.

Subjects then participated in the main experiment that was conducted in two parts. Subjects first played 10 periods of their assigned treatment, being matched with a different partner on the other side of the market in each period. At the start of period 11, new instructions were distributed concerning the change in information structure between treatments, which were read aloud. Subjects then played 10 additional periods, again matching with the same partner at most once.

Following a short questionnaire in which gender and other demographic information were recorded, payments to the subjects were made in cash based on the earnings they accumulated throughout the experiment with an exchange rate of 10 ECU to $1 AUD. In addition, each subject received a show-up fee of $10. Since payoffs during the experiment could be negative, the subjects could use the show-up fee to prevent bankruptcy during the experiment.[16] The average payment at the end of the experiment was \$51.10 AUD.[17] At the time of the 2009 experiments \$1 AUD = \$0.80 USD.

3.3. Hypotheses

The Moore–Repullo mechanism used in our experiment is designed to implement truthful announcements and efficient trade. Our predictions in the no-noise treatment are as follows:

HYPOTHESIS 1. *In the no-noise treatment buyers truthfully announce their signals and sellers do not challenge these announcements.*

Hypothesis 1 is based on three conditions that must be satisfied in order for the mechanism to function: the counter-offer condition, the appropriate-challenge condition and the truth-telling condition. Each of these conditions has implicit assumptions about how individuals behave and require at least some consistency between an individual's beliefs and the actions of other individuals at later stages of the game.

The counter-offer condition requires that a buyer who is appropriately challenged is willing to accept the counter-offer instead of rejecting it. As discussed in detail in Fehr et al. (2014), there is strong evidence of nonpecuniary benefits for rejecting an appropriate challenge when individuals are negatively reciprocal. In the current paper, we wanted to focus on the impact of imperfect information, irrational beliefs, and strategic uncertainty and chose parameters where the counter-offer condition was likely to be met even when buyers were negatively reciprocal.[18]

The appropriate-challenge condition requires that sellers make appropriate challenges but not inappropriate challenges. In order to maximize the incentive of sellers to make appropriate challenges across a large range of potential beliefs about the buyer's behavior at the counter offer stage, we chose to pass the fine F to the seller in the case that the counter-offer is accepted.

Finally, for the truth-telling condition to hold, it must be that a buyer, given his beliefs about the actions of the seller, has an incentive to make a truthful announcement

[16] Although we had no bankruptcies in the experiment, there is a potential that the description of bankruptcy rules could prime individuals to be more loss averse in the experiment. To check for this, the eight additional control treatments without beliefs paid only for a single period and increased the show-up fee to \$35 to cover the worst outcome. We find no significant difference in our results.

[17] The experiment took roughly 1.5 h resulting in an hourly pay rate of \$34.07 AUD. The Australian minimum wage is \$21.08 for casual employment.

[18] In our no-noise treatment, a buyer with a high signal who retaliates after a low announcement and a challenge must prefer the payoffs of $\{-25, -25\}$ for the Buyer and Seller over payoffs of $\{20, 50\}$. Equivalently, he must be willing to destroy \$0.60 of his own money to destroy \$1.00 of the seller's money after a low announcement and a challenge. This is much larger than what is seen in standard ultimatum games. For example, in a \$10 ultimatum game such a high level of required reciprocity implies that an offer of \$3.75 is rejected—an event that almost never occurs in subject pools such as ours.

rather than a lie. This decision is based on the buyer's belief about the likelihood of being challenged after a lie and the likelihood of being challenged after a truthful announcement.

As the truth-telling condition is based on the incentives generated from the lower two stages, we parameterized the experiment with an eye toward making each of the intermediate conditions as slack as possible. In places where parameters affected multiple constraints simultaneously (such as the fine size or counter-offer price) we erred toward ensuring that the counter-offer condition was satisfied. We also set the prices in the absence of a challenge equal to half of the buyer's announcement in order to minimize the importance of fairness considerations and to make the subgame perfect equilibrium salient.

3.3.1. The Noise Treatments. As soon as one introduces noise in agents' information about the state of nature (i.e., about the value of the good to be traded), the truth-telling equilibrium vanishes and pure and mixed strategy equilibria arise in which either: (i) the buyer makes announcements that are different to his signal; and/or (ii) the seller challenges announcements that are the same as her signal.

In Appendix A of the paper, we show that our environment generates two pure strategy equilibrium and a mixed strategy equilibrium when noise is introduced. In the discussion below, we concentrate on the mixed strategy equilibrium because the pure strategy equilibrium are based on out-of-equilibrium beliefs that are unlikely to exist in our experimental environment.[19]

Taking into consideration only pecuniary incentives, the mixed strategy equilibrium in our game involves buyers truthfully announcing their signals and sellers who mix between challenging and not challenging when they observe a low signal and a low announcement. Buyers in this equilibrium mix between accepting and rejecting the counter offer when they are challenged after making a low announcement with a low signal.

Unlike our no-noise environment where we can minimize the impact of negative reciprocity on the equilibrium predictions of the game, the mixed strategy above requires that the buyer is indifferent between accepting and rejecting the counter-offer when he is challenged. When buyers are negatively reciprocal and view challenges of the seller as an unkind act, they may be willing to reduce their own expected payoffs in order to reduce the earnings of a seller who has challenged them. Thus, even small amounts of negative reciprocity will change the behavioral properties of the mixed strategy equilibrium. We describe this in more detail in Appendix A of our paper. Here

[19] In the pure strategy equilibrium that involves buyer lies, the buyers' lies are supported by an out-of-equilibrium belief of sellers that any challenge they make will be rejected. Given that we maximized the incentives of buyers to accept challenges, we did not expect that sellers would hold these beliefs and did not expect them to be common knowledge. In the other pure-strategy equilibrium, buyers always announce high because they fear that sellers will always challenge a low announcement even when they observe a low signal. Although such beliefs may be more plausible, we do not find evidence of such buyer strategies in the empirical data.

we only provide a short summary of the impact of negative reciprocity on the mixed strategy equilibrium.

If buyers are moderately reciprocal, it will be the case that sellers no longer have an incentive to make false challenges (i.e., challenging after they received a low signal and the buyer announce low) and therefore an alternative mixed strategy equilibrium emerges in which buyers lie with positive probability and sellers mix between challenging and not challenging when they observe a high signal and are faced with a low announcement.

Structural estimates of reciprocity using data from Fehr et al. (2014) indicate that subjects are willing to sacrifice between \$0.17 and \$0.46 to destroy \$1 of wealth of a seller after a legitimate challenge in a related subgame-perfect implementation mechanism. Across this range of reciprocal preferences we expect to see mixed strategies with buyer lies and may also see mixed strategies with seller false challenges. Although different distributions of reciprocity will lead to slightly different point predictions, a property of the model is that the total lies by buyers and false challenges by sellers increases when noise is introduced. In addition, the challenges of buyers' lies and the rejection of false challenges is lower with noise than without noise. We summarize these prediction in the following hypotheses.

HYPOTHESIS 2. *The likelihood that a buyer with a high signal announces a low valuation is higher in the treatments with imperfect information. The likelihood that a seller with a high signal challenges a low announcement is lower in the treatments with imperfect information.*

HYPOTHESIS 3. *The likelihood that a seller with a low signal challenges a low announcement is higher in the treatments with imperfect information. The likelihood that a buyer accepts such a challenge although he received a low signal is also higher in the imperfect information treatments.*

4. Experimental Results

We describe the results of the experiment in this section. Section 4.1 uses the data from the no-noise treatments to study Hypothesis 1. Section 4.2 uses data on beliefs and from a number of additional experiments to interpret some of the results from Section 4.1. Section 4.3 uses data from both the no-noise and noise treatments to study Hypotheses 2 and 3.

We call a draw of a red ball the *high signal*, a draw of a blue ball the *low signal*, an announcement of 70 a *high announcement* and an announcement of 20 a *low announcement*. As before, we define a *lie* as an announcement by B of a low value after observing a high signal. We define an *appropriate challenge* as a challenge by S of a low announcement with the high signal, an *inappropriate challenge* as a challenge by S of a high announcement with the high signal, and a *false challenge* as a challenge by S of a low announcement with the low signal.

4.1. The Mechanism Under Perfect Information

Under Hypothesis 1, our experimental design predicts that in the no-noise treatment, the counter-offer condition, appropriate-challenge condition, and truth-telling condition will hold. These conditions imply that *B* will always tell the truth, *S* will make only appropriate challenges, and *B* will accept counter-offers if and only if they result from an appropriate challenge. The data from the no-noise treatment provides support for only two of these conditions.

RESULT 1. *The mechanism fails to induce truth-telling in a substantial number of cases. This occurs despite the fact that sellers appropriately challenge buyers' lies most of the time and buyers accept these (appropriate) challenges and reject false challenges most of the time.*

Figure 2 displays the patterns of play we observed in the no-noise treatment of the experiment. The left column examines play when an individual receives a low signal whereas the right side examines play when an individual receives a high signal. Panel (a) summarizes *B*'s announcement decision, Panel (b) summarizes *S*'s challenge decision, and Panel (c) summarizes *B*'s decision to accept or reject counter-offers. The error bars show 95% confidence intervals of each proportion with standard errors clustered at the individual level.

Panel (a) shows that after a low signal, 97.2% of individuals announce that the value is low. By contrast, after a high signal, 30.8% deviate from the theoretical prediction of Hypothesis 1 and lie. We discuss this deviation from truth-telling in greater detail below after detailing play in the other stages of the game.

Panel (b) shows the proportion of announcements that are challenged after each combination of announcement and signal. As can be seen, a low announcement with a low signal is challenged only 4.8% of the time whereas a high announcement with a high signal is challenged only 4.1% of the time. This implies that inappropriate challenges rarely occur in the data. By contrast, *S*s challenge a low announcement with a high signal 93.4% of the time implying that *S*s almost always make appropriate challenges.

Finally, Panel (c) shows the proportion of counter-offers that are accepted for each combination of announcement and signal. In the case of a high signal, *B*s always reject counter-offers after truthful announcements and almost always accept counter-offers after a lie. In the case of a low signal, *B*s always reject challenges after a low announcement.

Although there are small deviations from the theoretical predictions of the model in the challenge stage and counter-offer stage, these deviations tend to vanish over time. Panel (a) of Figure 3 tracks the proportion of truthful announcements that are challenged in each period. This data is overlayed with the predictions and 95% confidence intervals from a simple linear random effects regression that regresses the challenge decision on the period. As can be seen, challenges of truthful announcements are diminishing and the proportion of truthful announcements that are challenged is not significantly different from the theoretical prediction of 0% by period 10. Similarly, as

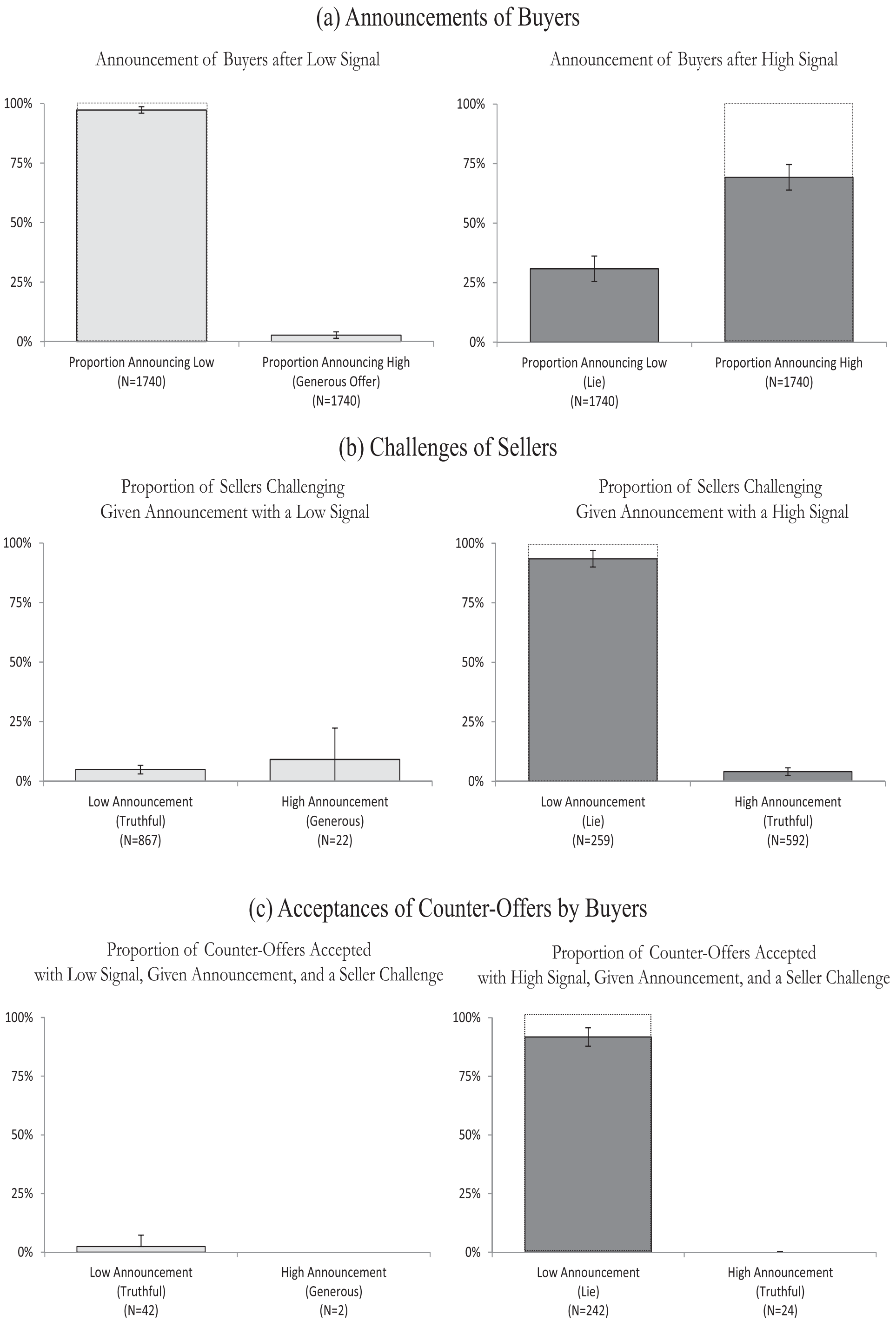
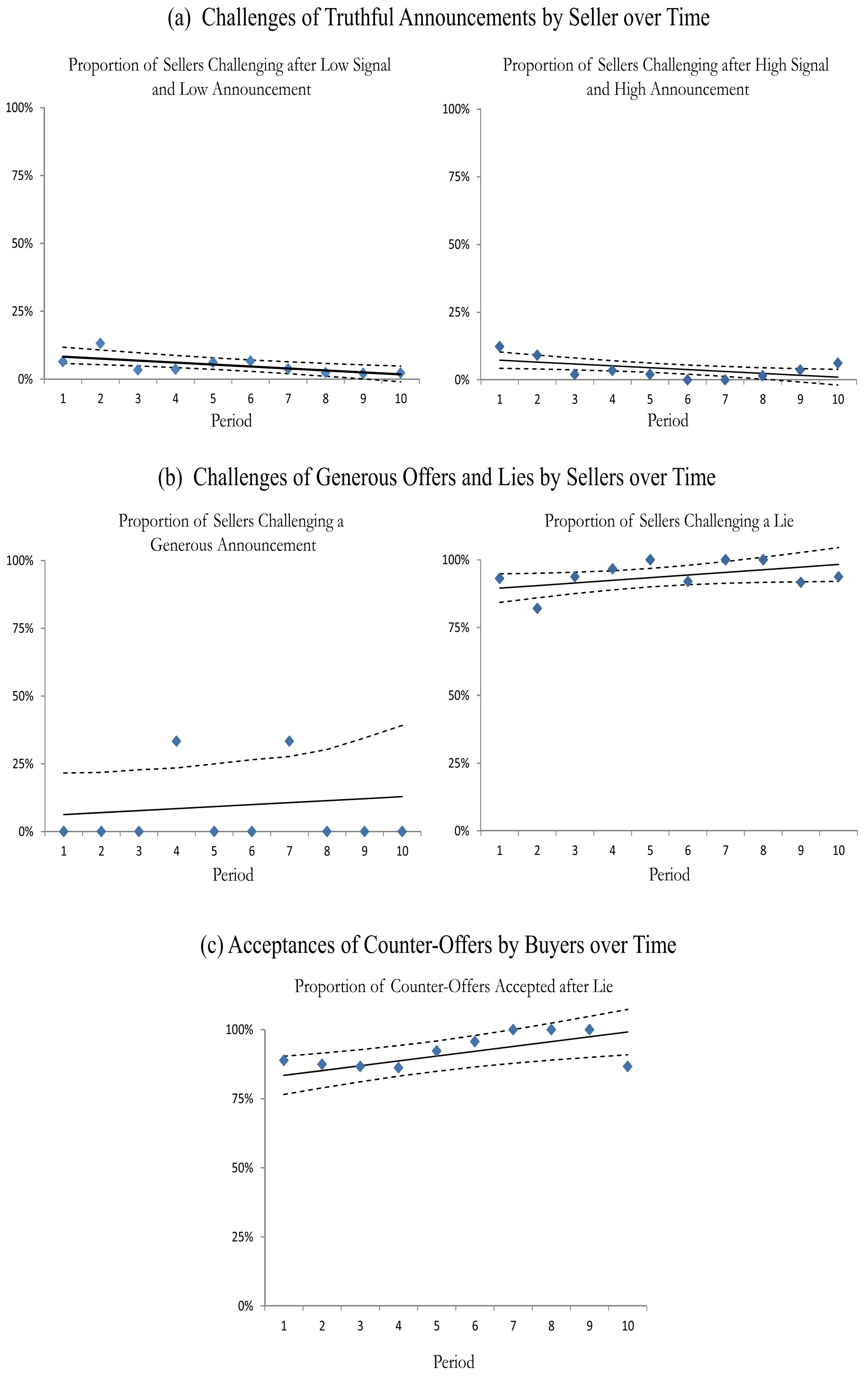


FIGURE 2. Pattern of play in no-noise treatment.



(a) Challenges of Truthful Announcements by Seller over Time

(b) Challenges of Generous Offers and Lies by Sellers over Time

(c) Acceptances of Counter-Offers by Buyers over Time

FIGURE 3. Evolution of play in challenge stage and counter-offer stage of no-noise treatment.

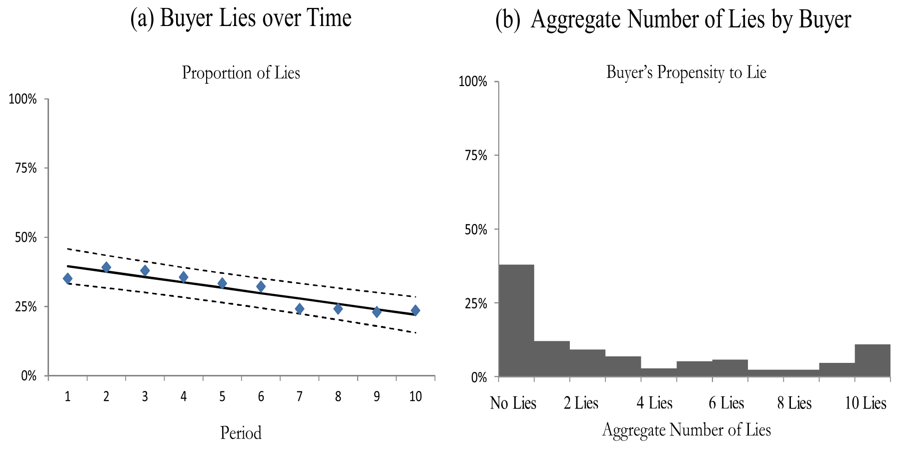


FIGURE 4. Evolution and distribution of lies in announcement stage of no-noise treatment.

seen on the right side of Panel (b), challenges of lies are increasing over time and the proportion of lies is not significantly different from the theoretical prediction of 100% by period 10. Taken together, the data strongly supports the appropriate-challenge condition.

Panel (c) of Figure 3 tracks the proportion of counter-offers that are accepted after a lie over time using the same construction of the prediction line and 95% confidence intervals as in the previous panels. Although some *B*s initially reject counter-offers, the proportion of counter-offers being accepted increases over time and is not significantly different to the theoretical prediction by period 10. Thus, there is strong evidence that the counter-offer condition is met in the data.

Given that the appropriate-challenge condition and the counter-offer condition hold, *B*s have pecuniary incentives to announce truthfully by construction of the mechanism. Thus, we might expect that lies—defined as low announcements with the high signal—converge to zero over time. Figure 4 shows that this is not the case. As can be seen in Panel (a), the proportion of *B*s who are lying is indeed slightly decreasing over time. However, this proportion is above 20% and significantly different from the theoretical prediction of 0% even in period 10. In fact, looking at the last four periods the rate of lying is constant at roughly 25%.[20]

Panel (b) shows a histogram of *B*'s lie rates in the no-noise treatment using all periods. As can be seen, 37.9% of *B*s never lie in the no-noise treatment whereas

[20] The rate of lying is 23.9% in the last four periods of sessions that begin in the no-noise treatment and 23.5% in sessions that ended with the no-noise treatment. Thus the ordering of treatments does not appear to influence the rate of lying. As discussed in section 4.2, we also ran 30 and 40 period follow-up sessions of the no-noise treatment where we elicited incentive compatible beliefs. The rate of lying in the last 5 periods of these sessions was close to 40% and there is no evidence of convergence toward zero.

10.9% of individuals lie in every period. This bimodal distribution becomes more pronounced over time: in a restricted sample of the last five periods of the treatment, 60.9% of *B*s never lie whereas 17.2% lie in each period. Thus, while many individuals stop lying over time a significant subset of individuals do not stop lying. We explore why these individuals may find it in their interest to lie in the next section.

4.2. Understanding Deviations from Truth-Telling in the No-Noise Treatment

One potential reason for the failure of subgame-perfect implementation is that individuals must place a large amount of faith in the rationality of other players. *B*s who announce truthfully must have faith that *S*s will not make an inappropriate challenge. However, if a *B*'s fear of such an inappropriate challenge is high enough, it may be in his best interest to adopt a strategy that minimizes his potential losses.

In practice, it is relatively rare for *S*s to make an inappropriate challenge. Nonetheless, the belief that some *S*s challenge a truthful high announcement may induce *B*s to lie. The implemented mechanism implies that a challenged high announcement will lead to relatively large losses for *B* regardless of whether *B* accepts or rejects the challenge. If *B* accepts the challenge, he will earn $70 - 75 - 25 = -30$; if he rejects the challenge, he will earn -25. These losses contrast sharply with the payoff of 20 that *B* can guarantee himself by lying, being challenged by *S*, and accepting the counter-offer.

Looking at the beliefs data of *B*, it appears that the fear of inappropriate challenges is strongly correlated with lies. Table 2 reports the results of regression analysis where the dependent variable is 1 if *B* lies after the high signal and 0 if *B* makes a truthful announcement. This variable is regressed on the belief that a lie will be challenged and the belief that a truthful announcement will be challenged. To allow for potential nonlinearities in the beliefs data we treat *B*'s beliefs as categorical data and split the 4-point Likert scale into a series of dummy variables. We use the category "Never" as the omitted dummy category. Column (1) reports the results of a simple linear probability model with errors clustered at the individual level. Column (2) reports the results of a fixed effects regression with both time and individual level fixed effects.

As can be seen in column (1), *B*'s belief about the likelihood that he will be challenged after a truthful announcement correlates with his likelihood of making a lie. *B*s are 39.7 (59.0) percentage points more likely to lie if they believe that a truthful announcement is "Likely" ("Always") to be challenged relative to an individual who believes a truthful announcement will "Never" be challenged. The probability of making a lie is increasing as an individual's beliefs becomes more pessimistic suggesting a monotonic relationship between beliefs and lies. This conclusion also holds if we control for individual and time fixed effects (see column (2)).

4.2.1. Precise Quantification of Beliefs to Better Understand Buyer Lies. To explore further the way in which beliefs may be guiding lies in the no-noise treatment we ran an additional experiment in which we elicited probabilistic beliefs of being challenged using an incentive-compatible elicitation mechanism developed in Savage (1971) that

TABLE 2. Linear probability model of lies by buyers.

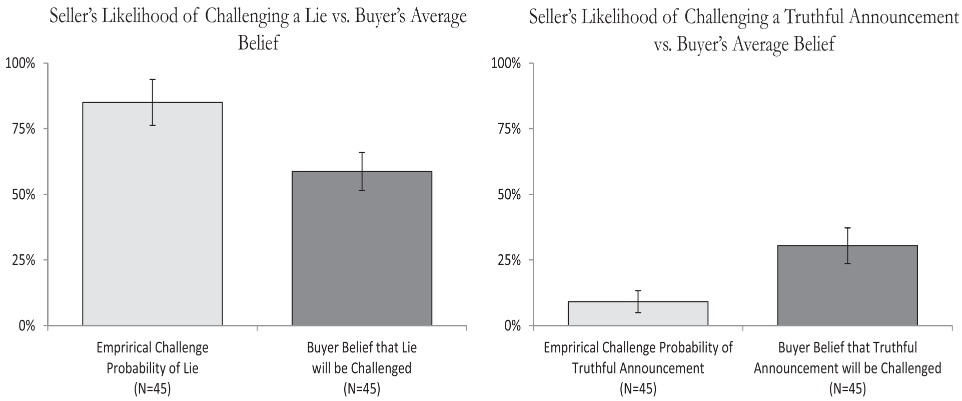
Buyers belief that seller will challenge a high announcement with high signal	(1)	(2)
"Unlikely"	0.065	0.025
	(0.051)	(0.044)
"Likely"	0.397***	0.186***
	(0.070)	(0.055)
"Always"	0.590***	0.234***
	(0.074)	(0.063)
Buyers belief that seller will challenge a low announcement with high signal		
"Unlikely"	– 0.027	– 0.170***
	(0.089)	(0.064)
"Likely"	– 0.040	– 0.024
	(0.071)	(0.064)
"Always"	– 0.127*	– 0.113*
	(0.066)	(0.059)
Constant	0.249***	0.325***
	(0.060)	(0.049)
Individual fixed effects	No	Yes
Time fixed effects	No	Yes
R^2	0.203	0.156
Observations	851	851

Dependent variable is 1 if the buyer lies by announcing low with a high signal and 0 otherwise. The omitted category is Seller "Never"Challenges. Regression (1) is a linear probability model with errors clustered by individual. Regression (2) is a fixed effect regression with both time and individual fixed effects. *Significant at 10%; **Significant at 5%; ***Significant at 1%.

is shown by Karni (2009) to induce truthful reporting of beliefs for rational agents with any von Neumann–Morgenstern utility function.[21] In this follow-up treatment, we restricted attention to only the no-noise treatment and ran additional periods to study convergence. We ran two sessions with 30 periods and two sessions with 40 periods with random matching across periods. A total of 90 individuals participated in the experiment. The details of this elicitation mechanism can be found in Appendix B.[22]

[21] Akin to a standard BDM mechanism (Becker, DeGroot, and Marschak 1964), the belief elicitation mechanism gives B a dominant strategy to announce his true beliefs by using B's reported belief to assign him to one of two lotteries—one that is contingent on S's challenge decision and one that is independent of this decision—across a set of binary lottery pairs. We randomly select one of these lottery pairs to be played so that beliefs impact the assignment of B to a lottery but not the explicit characteristics of this lottery. We use the strategy method in this follow up experiment for S's challenge decisions as we want to elicit incentive-compatible beliefs from B about the likelihood of being challenged after a truthful announcement and after a lie. To do so we need to know S's challenge decision for both announcements. See Appendix B for full details.

[22] As we were concerned with potential hedging, the follow-up experiment paid only for one period of the experiment and only for the announcement game or the belief elicitation game. There was a 50% chance that the announcement game would be paid and a 50% chance that one announcement-signal combination of the belief elicitation game would be paid.


Seller's Likelihood of Challenging a Lie vs. Buyer's Average Belief

Seller's Likelihood of Challenging a Truthful Announcement vs. Buyer's Average Belief

FIGURE 5. Buyer beliefs about the probability of challenge relative to empirical challenge probabilities of sellers.

RESULT 2. *The majority of Bs have pessimistic beliefs about being challenged after a truthful announcement of 70. The majority of Bs have optimistic beliefs about being challenged after a lie of 20.*

Figure 5 compares the empirical challenge probability of S's to B's belief of being challenged. Both the means and 95% confidence intervals shown are calculated from individual averages. As can be seen on the right hand side of the figure, buyers are strongly pessimistic about the likelihood of being challenged after a truthful high announcement. Although the empirical probability of being challenged is 9.1%, the average belief is 30.4%. This pessimism is prevalent across the population, with 80.1% of individuals having pessimistic beliefs about being challenged relative to the empirical distribution. The difference of beliefs and the empirical distribution is significant in both a simple t-test ($t = -5.38$, p-value < 0.01) and a Mann–Whitney–Wilcoxen test ($z = -5.13$, p-value < 0.01).[23]

Vice versa, buyers are optimistic about the likelihood of being challenged after a lie with a high signal. Although S's challenge 85.0% of the time after a lie (a 15.0% deviation from the Nash Equilibrium), the average belief is 58.7% (a 41.3% deviation from the Nash Equilibrium). This optimism is again prevalent across the population, with 76.7% of individuals having optimistic beliefs about being challenged relative to the empirical distribution. The difference between beliefs and the empirical distribution is again significant (t-test: $t = 4.70$, p-value < 0.01; Mann–Whitney–Wilcoxen test: $z = 5.56$, p-value < 0.01).

Given the optimistic beliefs about outcomes after a lie and pessimistic beliefs about outcomes after truthful announcing, a natural hypothesis is that Bs may believe that they are monetarily better off lying than telling the truth. To test this hypothesis, we

[23] Observations are an individual buyer's average belief and an individual S 's average challenge rate over all periods.

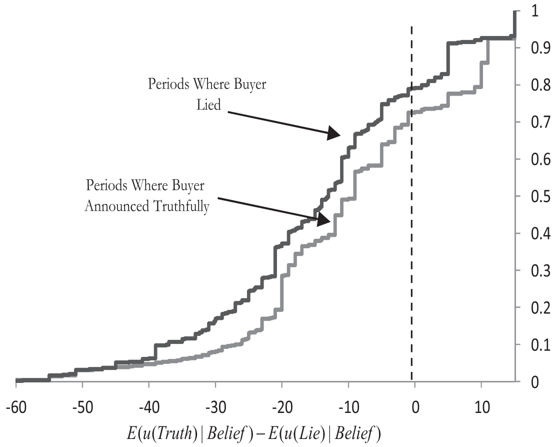


FIGURE 6. Cumulative distribution function of expected gain from telling the truth relative to lying split between observations where the buyer is lying (dark gray) and telling the truth (light gray).

use B's reported beliefs to compute the expected value of lying and telling the truth after a high signal if Bs respond optimally to a subsequent challenge. We next take the difference between these expected values to estimate the expected monetary gain from truth-telling.

RESULT 3. *The majority of Bs believe they have a higher expected value from lying compared to truth-telling after a high signal. Bs with more optimistic beliefs about being challenged after a lie and more pessimistic beliefs about being challenged after a truthful announcement are more likely to lie.*

Figure 6 show the empirical cumulative distribution functions (CDF) of the expected gain from truth-telling split between observations where an individuals is lying ($N = 543$) and observations where an individual is telling the truth ($N = 491$).[24] As can be seen, the empirical CDF of the expected monetary gain from truth-telling for individuals who tell the truth first order stochastically dominates the CDF for individuals who lie, suggesting that heterogeneity in beliefs is an important factor in the decision to announce truthfully.[25] For both distributions, however, the proportion of individuals where the expected monetary gain from truth-telling is negative is large,

[24] We restrict attention to observations where (i) the buyer believed that announcing low with a high signal had a higher chance of being challenged than announcing high with a low signal and (ii) the buyer announced low with a low signal. There is a very small fraction of individuals in our sample that do not satisfy these plausibility conditions. If we include them, all the qualitative results remain the same and significant.

[25] These distributions are significantly different in a bootstrapped version of the Mann–Whitney–Wilcoxen test where we randomly sampled a single period from each buyer in each iteration. p-value $<$ 0.01. See Datta and Satten (2005) for a discussion.

with 79.2% (72.7%) of observations where the buyer lies (tells the truth) falling into this category.[26]

One potential reason for the high level of pessimism seen in *B*'s beliefs about being inappropriately challenged is that at least a subset of individuals are choosing announcement strategies that limit their ability to learn over time. 26.7% of individuals lie in each of the last 10 periods of the session and in at least 90% of periods overall. These individuals account for 64.2% of overall lies and 66.7% of lies that occur in the last 10 periods. As a *B* who lies in each period gets no new information about the likelihood of being challenged after a truthful announcement, her actions inhibits her ability to learn.[27]

Overall, our data suggests something of a paradox in the functioning of the Moore–Repullo mechanism. Although the mechanism was designed to induce truth-telling using pecuniary incentives, most individuals who are truthful are distrustful of their partner and believe that such actions will lead to monetary loss. Truthful announcements are therefore being supported not by pecuniary incentives, but instead by nonpecuniary ones.

4.3. The Mechanism Under Almost-Perfect Information

The theoretical discussion in Section 2 predicts that as we introduce imperfect information about the value of the good, additional breakdowns in the mechanism will occur. As described in Hypothesis 2, *B*s with high signals are predicted to lie with greater frequency and *S*s are predicted to reject the buyers' lies with lower frequency. Further, as described in Hypothesis 3, *S*s are predicted to challenge low announcements although they received low signals (what we call a false challenge) and *B* 's are predicted to accept some of these false challenges. We find support for most of these theoretical predictions:

RESULT 4. *The introduction of noise leads to a significant increase in B's lies and a small and weakly significant decrease in challenges of low announcements by Ss with a high signal. In addition, the introduction of noise also increases B's belief that even*

[26] Recent work by Holt and Smith (2016) studies a belief elicitation procedure similar to the one we use here and does not find any systematic bias in reports toward the mean. Burford and Wilkening (2016) studies the exact mechanism used here and also does not find any systematic bias in reports. Nonetheless, because empirically accurate beliefs lie on the boundary of the interval, we note that our calculation here may be sensitive to measurement error.

[27] The implementation mechanism that we study is based on the auxiliary assumption that individuals concentrate on the subgame-perfect equilibrium of the game. However, there exists other refinement concepts that may better describe behavior in our environment. One such refinement is the set of consistent self-confirming equilibrium, which requires that each player correctly predicts play at all information sets that can be reached when the player's opponents, but not the player herself, deviate from their equilibrium strategies. In the mechanism we study, the set of consistent self-confirming equilibrium includes the subgame perfect equilibrium and a second equilibrium where a buyer with a high signal lies, is challenged, and accepts the counteroffer. Buyers who play according to this second self-confirming equilibrium do not observe the sellers' challenge decisions after a truthful announcement and cannot update their initial beliefs. See Fudenberg etal. (1988), Fudenberg and Levine (1993), and Kalai and Lehrer (1993) for a discussion of this alternative refinement and its relation to learning.

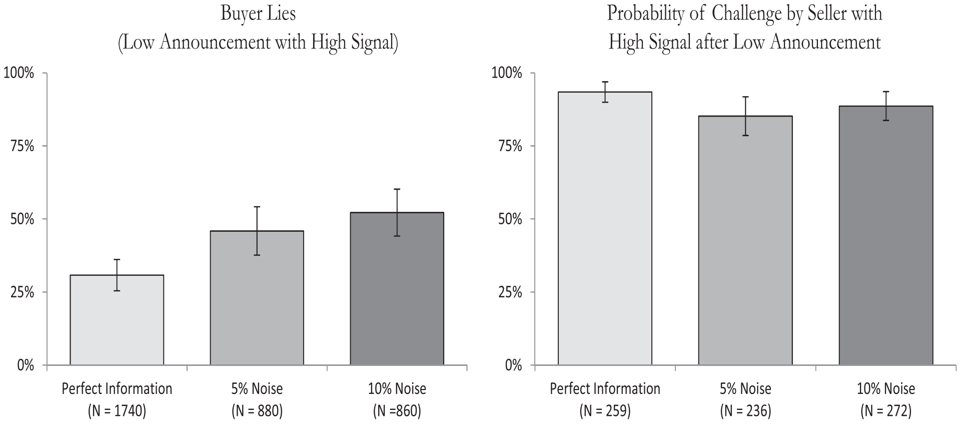


FIGURE 7. Buyer lies and seller's probability of challenging with high signal after a low announcement.

truthful announcements of a high signal will be challenged. Finally, noise also leads to an increase in both false challenges by S's and B's acceptance of challenges with a low signal after a low announcement.

An interesting aspect of Result 4 is that it confirms theoretically predicted behavioral tendencies that undermine the mechanism but, in addition, the evidence also shows that noise tends to exacerbate problems with the mechanism that we have already observed in the no-noise treatment: *B*s have more pessimistic beliefs that truthful announcements of a high signal will be challenged in the noise treatment compared to the no-noise treatments. This finding is in contrast to the theoretical prediction that truthful announcements of high signals should never be challenged in any of these treatments.

The left hand side of Figure 7 shows the proportion of *B*s with a high signal who lie across the three treatments. The error bars show 95% confidence intervals of each proportion with standard errors clustered at the individual level. As can be seen, *B*s lie in 45.9% of cases in the 5% noise treatment and in 52.2% of cases in the 10% noise treatment. Both of these lie rates are significantly higher than those in the no-noise treatment, where lies occur in 30.8% of cases (p-value < 0.01 in both treatment comparisons).[28]

The right hand side of Figure 7 shows that there is a small and weakly significant decrease in the challenges of low announcements with the high signal relative to the

[28] Using a single dummy for both noise treatments, the treatment effect is also significant at the 0.01 level when data is restricted to only the first or second treatment in a session (difference in first treatment: 0.20; difference in second treatment: 0.16) and when data is restricted to sessions that started in the no-noise treatment or the noise treatments (no-noise treatment first: 0.11; noise treatments first: 0.26). Other than a slightly higher lie rate in the early periods of the first treatment, there does not appear to be any order effects in our main experiment.

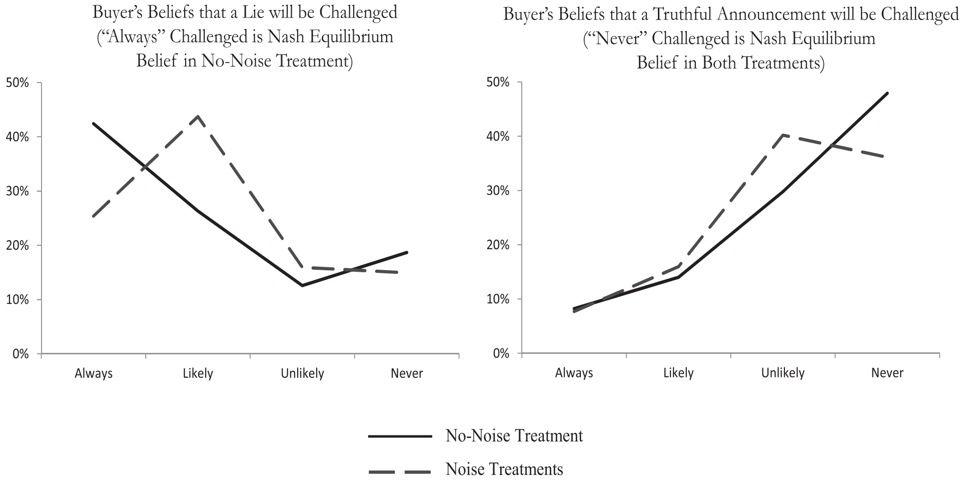


FIGURE 8. Buyer's beliefs after high signal.

no-noise treatment, where 93.4% of cases were challenged; in the 5% and 10% noise treatments the proportion of cases challenged were 85.2% and 88.6% respectively (5% noise treatment: p-value < 0.01; 10% noise treatments: p-value $= 0.087$). All three challenge rates are high, however, indicating that it would not be in B's interest to lie if their beliefs were consistent with the empirical challenge distributions of the sellers.

Our theoretical model predicts that the increase in lies in the noise treatment is driven by B's belief that S is less likely to challenge a lie. The left panel of Figure 8, which reports B 's belief that a lie will be challenged given a high signal, supports the existence of this channel. In the no-noise treatment 42.4% of individuals believe that a lie will always be challenged, whereas in the noise treatments only 25.4% of individuals hold this belief. Thus, in the noise treatments the buyers are much more optimistic that they can get away with a lie. This difference in beliefs across the noise treatments and the no-noise treatment is significant based on an ordered probit regression that regresses B's beliefs on a noise treatment dummy that is 0 for the no-noise treatment and 1 for the noise treatments ($z = -2.45$, p-value $= 0.014$, standard errors clustered by individual).[29]

The right panel of Figure 8 shows that the belief pattern observed in the no-noise treatment, namely that B's believe that even truthful announcements of a high value will be challenged, is exacerbated by the existence of noise. Although 48.1% of individuals believe that a truthful announcement will never be challenged in the no-noise treatment, only 36.1% of subjects in the noise treatments have this belief.

[29] Beliefs in this ordered probit regression and the one in the following paragraph are treated as categorical data with "Never" treated as category 0 and "Always" treated as category 3. The ordered probit has no other explanatory variables beyond the noise treatment dummy. The null hypothesis is that the noise treatment dummy coefficient is zero.

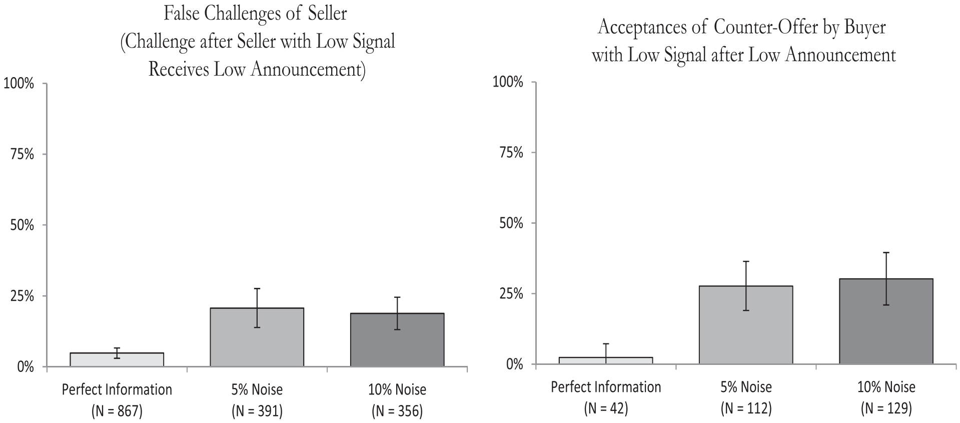


FIGURE 9. Seller's false challenges and buyer's probability of accepting a counter-offer with low signal after a low announcement.

This difference is significant in an ordered probit regression of buyers' beliefs on the same noise treatment dummy used above ($z = 2.21$, p -value $= 0.027$, standard errors clustered by individual).

Taken together, the belief pattern observed in Figure 8 suggests that there are two reasons why noise increases buyers' lying behavior. First, noise induces *B*s to believe that a lie is less likely to be challenged and, second, it strengthens the belief that truthful announcements will be challenged. Both reasons reduce the perceived pecuniary benefits from telling the truth relative to telling a lie. We study the causal impact of the second channel on buyers' announcement behavior in both the noise and the no-noise treatment in the next section.

Our results for the noise treatments also support the predictions of Hypothesis 3. Figure 9 shows the proportion of *S*s who make a false challenge and the proportion of *B*'s who accept a counter-offer after they received a low signal and announced a low value in each of the three treatments. The error bars show 95% confidence intervals of each proportion with standard errors clustered at the individual level. As can be seen on the left hand side, although there are very few false challenges in the no-noise treatment, the proportion of false challenges increases to 20.7% in the 5% noise treatment and 18.8% in the 10% noise treatment. Both noise treatments have significantly more false challenges than in their respective no-noise treatments based on a linear regression with errors clustered at the individual level (p-value < 0.01 in both cases).

As can be seen on the right hand side, *B*s are also much more likely to accept a counter-offer with a low signal and a low announcement in the noise treatments than in the no-noise treatment. Although B's accepted a challenge after a low announcement and a low signal in only 2.4% of observations in the no-noise treatment, they accepted 27.7% of such challenges in the 5% noise treatment and 30.2% of such challenges in the 10% noise treatment. Both noise treatments have significantly more acceptances

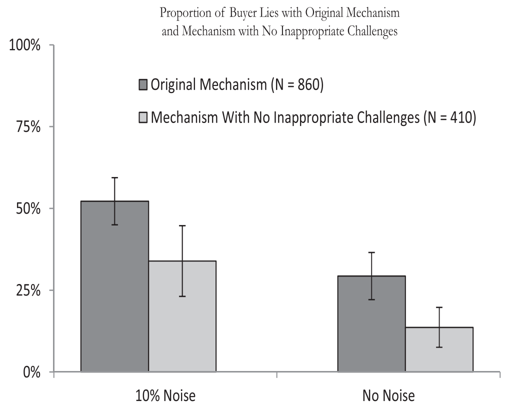


FIGURE 10. Frequency of buyer lies with original mechanism and alternative simple mechanism where high announcements cannot be challenged.

of counter-offers after a low announcement and a low signal than their respective no-noise treatment based on a linear regression with errors clustered at the individual level (p-value < 0.01 in both cases).

4.4. Robustness Checks

4.4.1. Does the Fear of Inappropriate Challenges Induce Buyers' to Lie?. If the belief that truthful announcements will be challenged is the main driver of lies in the no-noise treatment and also drives a subset of lies in the noise treatment, then eliminating the potential of such challenges should increase the likelihood of truth-telling in both treatments. We test this hypothesis by running four additional sessions where we eliminated the ability for S to challenge a B who makes a high announcement. Two of the sessions started in the 10% noise treatment and ended in the no-noise treatment whereas in the other session, individuals started in the no-noise treatment and ended in the 10% noise treatment. This "no-inappropriate challenge" mechanism is expected to increase the expected gain from truth-telling in both the noise and the no-noise treatments. We expect, therefore, that a large proportion of lies will decrease in this treatment relative to the baseline but that a significant portion of the gap between the no-noise and noise treatments will remain. A total of 82 individuals participated in these additional experiments.

RESULT 5. *Eliminating the ability of S to challenge high announcements substantially reduces Bs lies in both the no-noise treatment and the noise treatment. The introduction of noise leads to an increase in B's lies in both the baseline mechanism and the new mechanism.*

Figure 10 shows the proportion of lies in the original sessions with 10% noise and the new sessions using the no-inappropriate challenge mechanism. The error bars show 95% confidence intervals with standard errors clustered at the individual level. As can be seen, lies in both the noise treatment and the no-noise treatment decrease with the no-inappropriate challenge mechanism as we would expect if pessimistic beliefs about being challenged after a truthful announcement is a major contributor to lying.[30] This decrease in lies is particularly pronounced when comparing the second treatment in each session, where buyer lies fell to only 7.1% in the no-noise treatment and 27.0% in the 10% noise treatment.

It is interesting to note that the type of sequential mechanism we tested in the above additional sessions is not capable of implementing all social choice functions. Moore (1992) calls mechanisms like this "simple sequential mechanisms" and provides conditions under which they can implement a desired social choice function. Roughly speaking, this requires that only one party has state dependent preferences, or that preferences are perfectly correlated.[31]

4.4.2. How Small is Small?. Although we chose the levels of 5% and 10% noise in order to have enough power to differentiate between treatments, AFHKT suggests that very small levels of noise can lead to a breakdown of the mechanism. To study whether deviations from perfect information impact the distribution of lies even for very small levels of noise, we ran four additional sessions where we started with 10 periods of a 1% noise treatment and ended with a no-noise treatment. A total of 82 individuals participated in these additional experiments. We compare this treatment to the sessions where we started with 10 periods of the 5% noise treatment and ended with a no-noise treatment.

RESULT 6. *Even a very small perturbation in common knowledge leads to an increase in lies relative to the no-noise treatment.*

Figure 11 shows the proportion of buyer lies and seller false challenges in the 5% noise treatment and 1% noise treatment with 95% confidence intervals clustered at the individual level. The dotted lines in each figure show the proportion of buyer lies and seller false challenges in the subsequent no-noise treatment.

As can be seen in the left hand panel, both the 5% noise sessions and 1% noise have significantly more lies in the noise treatment than in their corresponding no-noise treatment. The proportion of lies in the 5% and the 1% noise treatments is surprisingly similar; there is no significant difference in the proportion of buyer in these two treatments based on a linear regression where buyer lies are regressed on the treatment dummy for the 5% noise sessions ($t = 0.76$, p-value $= 0.450$).

[30] The difference in lie frequency between the original mechanism and the no false challenge mechanism is significant at the 10% level based on a Mann–Whitney test where the lie frequency of each individual is the variable of interest: $z = 3.21$, p-value < 0.01. Similar results hold for a linear regression with data clustered at the individual level (p-value < 0.01).

[31] See Nöldeke and Schmidt (1995) and Hoppe and Schmitz (2011) for work on simple "option contracts" that have promising properties in a one-sided hold-up environment even when renegotiation is possible. We deliberately explore the performance of a three-stage mechanism in our simple environment with one-sided hold-up, because if such mechanisms fail to work well in a simple environment, they are even more likely to fail in the more complex environments that necessitate their use.

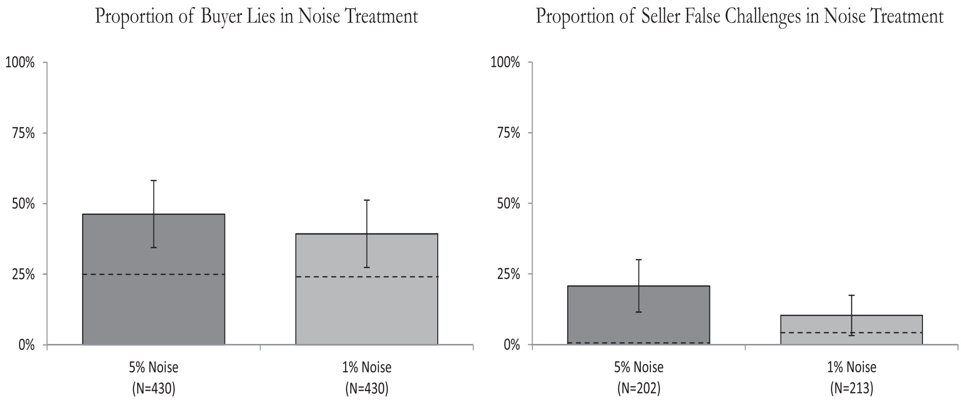


FIGURE 11. Proportion of lies and false challenges in the 5% noise treatment and 1% noise treatment.

As can be seen in the right hand panel, sellers make false challenges 10.3% of the time in the 1% noise treatment relative to 20.8% of the time in the 5% noise treatment—a difference that is weakly significant ($t = -1.91$, p-value $= 0.059$) based on a linear regression where sellers' false challenges are regressed on a dummy for the 5% noise treatment.

Taken together, although there is a small reduction in seller false challenges when noise rates decline, the large number of buyer lies in the 1% noise treatment illustrates that even small departures from common knowledge have a significant impact on the willingness of individuals to report truthfully. Our results thus illustrate the non-robustness of the Moore–Repullo mechanism to small amounts of noise.

5. Conclusion

In this paper we conducted a laboratory experiment to test the extent to which Moore and Repullo's subgame perfect implementation mechanism induces truth-telling in practice, both in a setting with perfect information and in a setting where buyers and sellers do not share common knowledge about the good's value. Our first finding is that even in the no-noise treatment, where no lies are predicted in equilibrium, buyers lie by announcing a low value with a high signal roughly 25% of the time. Our data suggests that in all treatments a substantial proportion of these lies are driven by pessimism about being inappropriately challenged after a high announcement. This pessimism is strong enough that a large majority of individuals who are telling the truth believe they would be better off lying, which suggests that the mechanism is being supported in part by nonpecuniary incentives for telling the truth.

Our second main finding is that the introduction of noise leads to an increase in buyers' lies and sellers' false challenges. The introduction of noise increases the proportion of buyers who announce a low value with a high signal by 15–25 percentage points; these lies are persistent and do not diminish with experience. Similarly, the proportion of sellers who falsely challenge in the noise treatments increases by 15 percentage points relative to the no-noise treatment. Lack of perfect information is behaviorally important even when the level of noise is reduced to a very small 1% level.

If we adjust the Moore–Repullo mechanism by ruling out false challenges, buyers' lying rate in the no-noise treatment decreases by 15.6 percentage points. Likewise, the institutional removal of such false challenges also decreases the lying rate in the noise treatments significantly. However, in the noise treatments this deviation from the Moore–Repullo mechanism does not solve the lying problem. Even if the fear of false challenges of high announcements is ruled out, a lying rate of 27% prevails in the 10% noise treatment, which indicates the pervasive influence of uncertainty regarding the good's value on lying behavior.

One important potential objection to our findings is that when parties themselves design a mechanism one should be less concerned about the fears of irrationality that play a prominent role in our experiment. As Eric Maskin suggested to us, when the parties are designing a contract they may engage in all sorts of discussion about how the game might be played. This is an important point and we believe that there is potential here for future theoretical and experimental research. From a theory standpoint, preplay communication can naturally be modeled as a cheap-talk stage prior to the mechanisms studied in this paper. To understand the benefits or the potential costs of such communication one should, and can, model this additional stage.[32] From an experimental standpoint, it would be interesting to see whether preplay communication makes buyers more confident about the rationality of their playing partner and whether improved confidence may lead to less buyer lies in environments without noise.

Our findings suggest several important avenues for future research, in addition to that mentioned in the preceding paragraph. First, the fact that individuals are willing to sacrifice their material well-being to tell the truth suggests that preferences for honesty should help implementation.[33] Second, in view of the empirical relevance of common knowledge, it also is important to design mechanisms that are robust to at least small amounts of imperfect information about the good's value. Third, it would be interesting to know (theoretically and empirically) how the introduction of asset ownership affects the functioning of extensive form mechanisms. In particular, asset ownership could be naturally modeled as an outside option for the asset holder, which in turn would affect either party's incentive to report the good's value truthfully or to challenge the other party. It would be interesting to see whether asset ownership helps achieve better equilibrium outcomes that are also robust to introducing small amounts

[32] Miettinen (2013) studies preplay communication in settings where agents may suffer guilt for reneging on their past promises. The author shows that when actions are strategic complements, informal agreements are easier to sustain than when actions are strategic substitutes.

[33] Current research by Kartik, Tercieux, and Holden (2014) suggests that when individuals have a known preference for honesty, full implementation can be achieved with simple mechanisms requiring only two rounds of iterated deletion of strictly dominated strategies.

of private information. Finally, similar experiments could be used to test the robustness of other implementation mechanisms, starting with virtual implementation. Overall, our analysis and findings in this paper raise a number of exciting issues to be tackled by future research.

Appendix A: Point Predictions of the Mixed Strategy Equilibrium

In this section, we derive the point predictions of the mixed strategy equilibrium of our game for each of our three noise treatments. We begin with the standard model where all participants are risk neutral and selfish. We then show how the point predictions of the model change when buyers receive positive utility for rejecting a challenge of the seller. This alternative model is discussed in detail in Fehr et al. (2014) where a sequential reciprocity equilibrium in the spirit of Dufwenberg and Kirchsteiger (2004) is developed.

A.1. The Mixed Strategy Equilibrium with Selfish Risk-Neutral Buyers

As in the main text, let the true valuation of the good be $\theta \in \{\theta^H = 70, \theta^L = 20\}$, with both states being equally likely. Let each player receive one of two possible signals, s^H and s^L, where s^H is a high signal correlated with θ being equal to 70, and where s^L is a low signal correlated with θ being equal to 20. Using the notation s_B^H (resp. s_B^L) to indicate that B received the high signal s^H (resp. the low signal s^L), the following table shows the joint probability distribution ν^ε over θ, the buyer's signal s_B, and the seller's signal s_S

ν^ε	s_B^H, s_S^H	s_B^H, s_S^L	s_B^L, s_S^H	s_B^L, s_S^L
$\theta = 70$	$\frac{1}{2}(1-\varepsilon)^2$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon^2$
$\theta = 20$	$\frac{1}{2}\varepsilon^2$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}(1-\varepsilon)^2$

For a given noise level ε, an action profile of a buyer consists of a probability of announcing low after observing each signal and a probability of rejecting the challenge given a signal and an announcement. Denote L^H as the probability of making a *low* announcement after observing a high signal and L^L as the probability of making a low announcement after a low signal. Further, let $R^{a_B|s_B}$ be the probability that the buyer rejects a challenge given his own announcement $a_B \in \{L, H\}$, his own signal $s_B = \{L, H\}$ and a challenge by the seller.

An action profile of the seller consists of a probability of challenging an announcement of the buyer for each potential announcement and signal. Let $C^{a_B|s_S}$ be the probability that the seller challenges given signal $s_S \in \{L, H\}$ and an observed announcement of the buyer $a_B = \{L, H\}$.

Although there are 10 potential mixing probabilities to specify in an equilibrium, we can use some of the structure of the mechanism to rule out mixing on some action sets. Let $P_{20} = 10$ and $P_{70} = 35$ be the trade prices without arbitration and let $P_A = 25$ and $P_B = 75$ be the counter-offer prices after announcing 20 and 70. A buyer who announces high and is challenged faces a price of $P_B = 75$ that is above his actual value of the good regardless of the state. Thus the buyer will always reject arbitration if he has announced high and $R^{H|L} = R^{H|H} = 1$. This also implies that the seller will never call the arbitrator if the buyer announces high, and thus $C^{H|L} = C^{H|H} = 0$. Further, a buyer who has a high signal and announces low will update his belief about the quality of the good based on the act of being challenged by the seller. However, for any equilibrium where the seller challenges with positive probability, the most pessimistic posterior the buyer can have after being challenged is that the state is low with probability $1/2$. (The posterior in the unlikely case where the seller challenges only with the low signal.) As the counter-offer price is 25 and the buyer's expected value for the good with this belief is 45, the buyer will always accept the counter-offer, and thus $R^{L|H} = 0$. Finally, the best a buyer can do with a low signal if he always announces high is to receive 35 with probability ε and -15 with probability $1 - \varepsilon$. If in equilibrium the buyer earns more than $35\varepsilon - 15(1 - \varepsilon)$ for a low announcement, it will be the case that $L^L = 1$.[34]

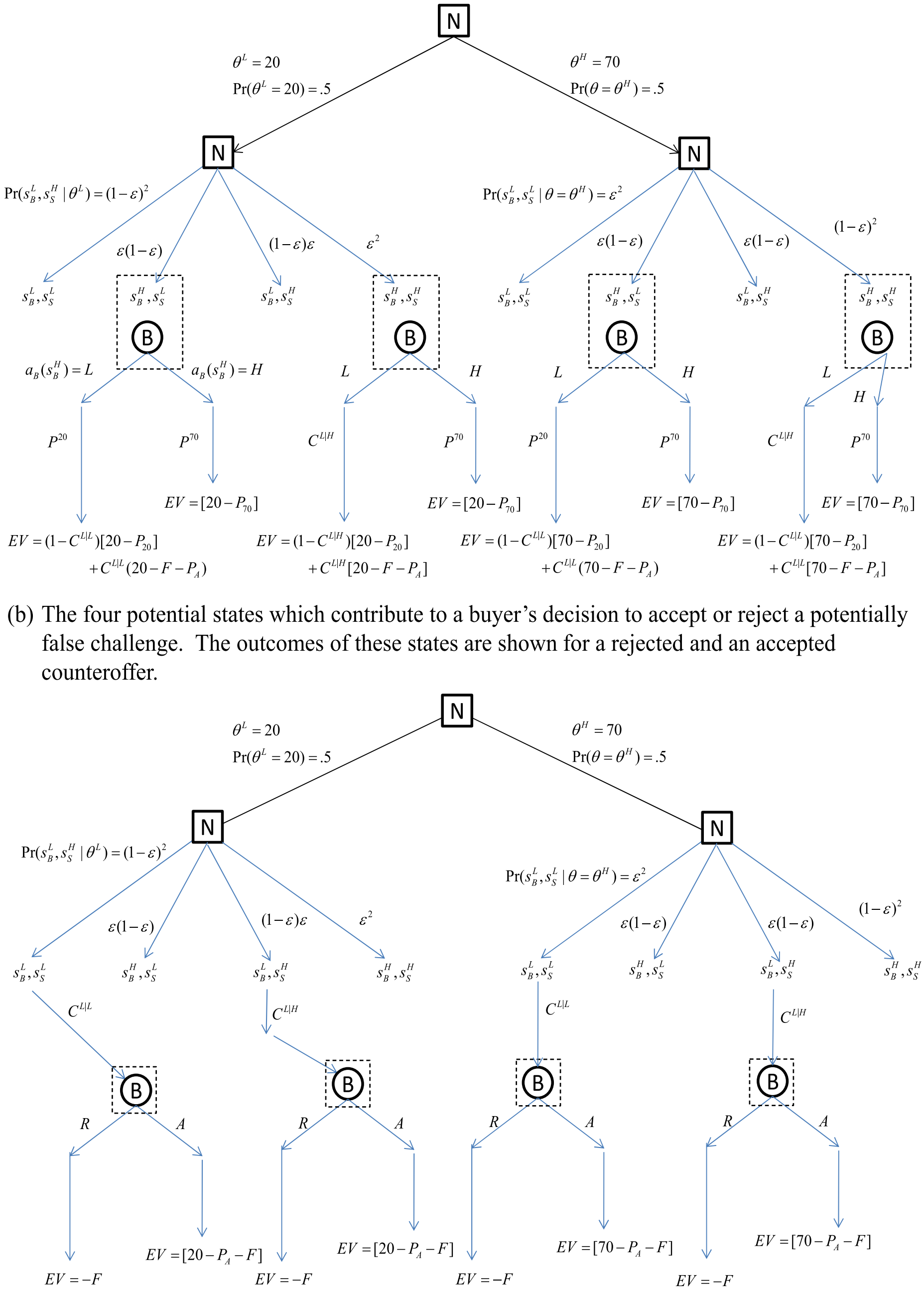
Taking as given the actions of buyers and sellers in the six states specified above, the mixed strategy equilibrium is based on (i) the proportion of times a buyer announces low given a high signal, L^H, (ii) the challenge probabilities given a low announcement, $C^{L|L}$ and $C^{L|H}$, and (iii) the probability that the buyer rejects a challenge given a low signal, a low announcement, and a challenge, $R^{L|L}$. These four mixing probabilities form the basis of all PBE where all stages of the subgame are reached and beliefs of both parties are consistent with the action profiles of the other party.

Given that beliefs of all parties must be consistent with their actions, a necessary condition for the mixed strategy equilibrium is that each individual is indifferent between each of their actions given the mixing probabilities of the other parties. These indifference conditions generate four linear constraints on the four mixing probabilities of the buyer and seller and generate a four-by-four linear system that derives unique point predictions. The construction of each linear constraint is as follows.

A.1.1. Buyer's Indifference Between Announcing Low and High with a High Signal. For the buyer to be indifferent between announcing high and low, the expected value of these announcements must be equal when aggregated over all potential states of nature.

Panel (a) of Figure A.1 shows the four potential states of nature where the buyer can have a high signal after nature draws the true value of the container and (conditional) signals for the buyer and seller. For each state, the expected value of each potential announcement is shown as a function of the challenge probabilities of the seller. For

[34] We argue in the main text that there is a pure strategy equilibrium where $L^L = 0$ and challenges never occur.


(a) The four potential states that contribute to the buyer's decision to lie by announcing low with a high signal. The outcomes of these states are shown for a low and a high announcement.

(b) The four potential states which contribute to a buyer's decision to accept or reject a potentially false challenge. The outcomes of these states are shown for a rejected and an accepted counteroffer.

N
$\theta^L = 20$
$\Pr(\theta^L = 20) = .5$
$\theta^H = 70$
$\Pr(\theta = \theta^H) = .5$
N
N
$\Pr(s_B^L, s_S^H \mid \theta^L) = (1-\varepsilon)^2$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)\varepsilon$
ε^2
$\Pr(s_B^L, s_S^L \mid \theta = \theta^H) = \varepsilon^2$
$\varepsilon(1-\varepsilon)$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)^2$
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
B
$a_B(s_B^H) = L$
$a_B(s_B^H) = H$
L
H
L
H
L
H
P^{20}
P^{70}
$C^{L|H}$
P^{70}
P^{20}
P^{70}
$C^{L|H}$
P^{70}
$EV = [20 - P_{70}]$
$EV = [20 - P_{70}]$
$EV = [70 - P_{70}]$
$EV = [70 - P_{70}]$
$EV = (1 - C^{L|L})[20 - P_{20}] + C^{L|L}(20 - F - P_A)$
$EV = (1 - C^{L|H})[20 - P_{20}] + C^{L|H}[20 - F - P_A]$
$EV = (1 - C^{L|L})[70 - P_{20}] + C^{L|L}(70 - F - P_A)$
$EV = (1 - C^{L|L})[70 - P_{20}] + C^{L|L}[70 - F - P_A]$
N
$\theta^L = 20$
$\Pr(\theta^L = 20) = .5$
$\theta^H = 70$
$\Pr(\theta = \theta^H) = .5$
N
N
$\Pr(s_B^L, s_S^H \mid \theta^L) = (1-\varepsilon)^2$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)\varepsilon$
ε^2
$\Pr(s_B^L, s_S^L \mid \theta = \theta^H) = \varepsilon^2$
$\varepsilon(1-\varepsilon)$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)^2$
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
$C^{L|L}$
$C^{L|H}$
$C^{L|L}$
$C^{L|H}$
B
R
A
R
A
R
A
R
A
$EV = -F$
$EV = [20 - P_A - F]$
$EV = -F$
$EV = [20 - P_A - F]$
$EV = -F$
$EV = [70 - P_A - F]$
$EV = -F$
$EV = [70 - P_A - F]$

FIGURE A.1. States contributing to the decision of the buyer to lie and reject a potentially false challenge.

example, as seen on the far left of the figure, with probability $\frac{1}{2}\varepsilon(1-\varepsilon)$, the buyer receives the high signal, the seller receives the low signal, and the true state of nature is low. If in this state the buyer announces low, he will not be challenged $1 - C^{L|L}$ percent of the time and be challenged $C^{L|L}$ percent of the time. As he has the high signal, he will always accept the counter-offer and thus these two outcomes yield values of $20 - P_{20} = 10$ and $20 - F - P_A = -30$ respectively. If, on the other hand, the buyer announces high, he will never be challenged (since $C^{H|L} = 0$) and receive $20 - P_{70} = -15$ for sure.

Taking into account the probability of each one of these potential states and the state's outcome, a buyer is indifferent between a high and low announcement if

$$\psi(\varepsilon)C^{L|H} + \delta(\varepsilon)C^{L|L} = \frac{P_{70} - P_{20}}{F + P_A - P_{20}}, \tag{A.1}$$

where $\psi(\varepsilon) = \varepsilon^2 + (1-\varepsilon)^2$ is the probability that the signals are the same for a given ε and $\delta(\varepsilon) = 2\varepsilon(1-\varepsilon)$ is the probability that they are different.

A.1.2. Buyer's Indifference Between Accepting and Rejecting a Challenge with a Low Signal and Low Announcement. In an equilibrium in which the seller is mixing between challenging and not challenging a low announcement with a low signal, it must be the case that the buyer is also indifferent between rejecting and accepting such a challenge. Panel (b) of Figure A.1 shows the probability of reaching this acceptance and rejection as a function of the signals and the challenge probabilities of the seller and under the assumption that $L^L = 1$. Taking into account the probability of each of these potential states and the state's outcome, a buyer is indifferent between rejecting and accepting the challenge if

$$C^{L|L} - \tau(\varepsilon)C^{L|H} = 0, \tag{A.2}$$

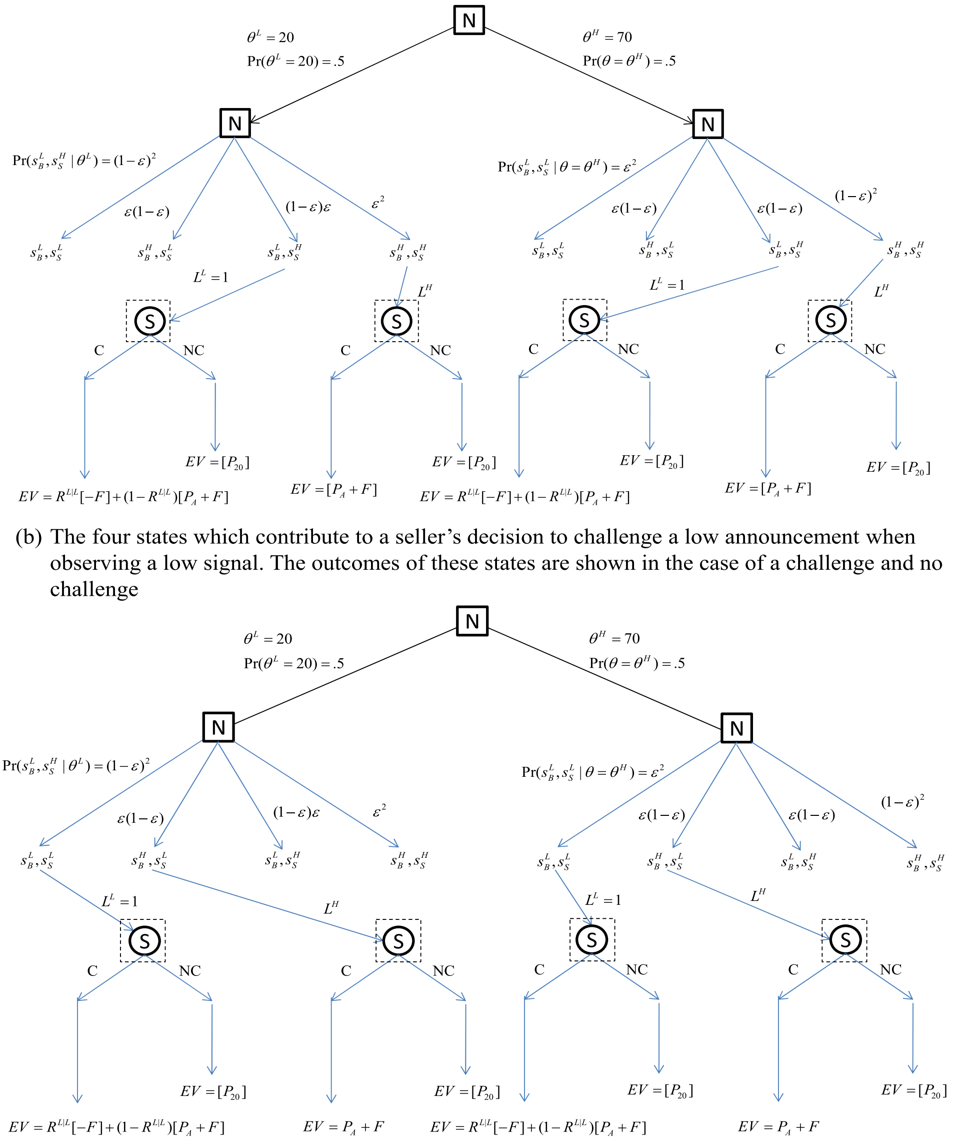
where

$$\tau(\varepsilon) = -\frac{\varepsilon(1-\varepsilon)[70 - P_A] + (1-\varepsilon)\epsilon[20 - P_A]}{\varepsilon^2[70 - P_A] + (1-\varepsilon)^2[20 - P_A]} \tag{A.3}$$

is the ratio of expected outcomes when the two parties have opposite signals relative to when they have the same signal. Note that $\tau(\varepsilon)$ is positive for all ε we consider since the denominator is negative.

A.1.3. Seller's Indifference Between Challenging and not Challenging After a Low Signal. As with the buyer, the seller's indifference for challenging after a low and high signal are based on the two mixing probabilities of the buyer. Panel (a) of Figure A.2 shows the expected value for challenging and not challenging for states of the world where the seller has a high signal and observes a low announcement. The likelihood of reaching each of these potential states is based on the likelihoods that the buyer will make a low announcement with each signal (L^H and $L^L = 1$) whereas the expected value of challenging is based on the likelihood that the buyer

(a) The four states which contribute to a seller's decision to challenge a low announcement when observing a high signal. The outcomes of these states are shown in the case of a challenge and no challenge


N
$\theta^L = 20$
$\Pr(\theta^L = 20) = .5$
$\theta^H = 70$
$\Pr(\theta = \theta^H) = .5$
N
N
$\Pr(s_B^L, s_S^H \mid \theta^L) = (1-\varepsilon)^2$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)\varepsilon$
ε^2
$\Pr(s_B^L, s_S^L \mid \theta = \theta^H) = \varepsilon^2$
$\varepsilon(1-\varepsilon)$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)^2$
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
$L^L = 1$
L^H
$L^L = 1$
L^H
S
S
S
S
C
NC
C
NC
C
NC
C
NC
$EV = [P_{20}]$
$EV = [P_{20}]$
$EV = [P_{20}]$
$EV = [P_{20}]$
$EV = R^{L|L}[-F] + (1-R^{L|L})[P_A + F]$
$EV = [P_A + F]$
$EV = R^{L|L}[-F] + (1-R^{L|L})[P_A + F]$
$EV = [P_A + F]$

(b) The four states which contribute to a seller's decision to challenge a low announcement when observing a low signal. The outcomes of these states are shown in the case of a challenge and no challenge

N
$\theta^L = 20$
$\Pr(\theta^L = 20) = .5$
$\theta^H = 70$
$\Pr(\theta = \theta^H) = .5$
N
N
$\Pr(s_B^L, s_S^H \mid \theta^L) = (1-\varepsilon)^2$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)\varepsilon$
ε^2
$\Pr(s_B^L, s_S^L \mid \theta = \theta^H) = \varepsilon^2$
$\varepsilon(1-\varepsilon)$
$\varepsilon(1-\varepsilon)$
$(1-\varepsilon)^2$
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
s_B^L, s_S^L
s_B^H, s_S^L
s_B^L, s_S^H
s_B^H, s_S^H
$L^L = 1$
L^H
$L^L = 1$
L^H
S
S
S
S
C
NC
C
NC
C
NC
C
NC
$EV = [P_{20}]$
$EV = [P_{20}]$
$EV = [P_{20}]$
$EV = [P_{20}]$
$EV = R^{L|L}[-F] + (1-R^{L|L})[P_A + F]$
$EV = P_A + F$
$EV = R^{L|L}[-F] + (1-R^{L|L})[P_A + F]$
$EV = P_A + F$

FIGURE A.2. States contributing to the decision of the seller to challenge with a high and low signal.

will accept this challenge ($R^{L|L}$ and $R^{L|H} = 1$). A seller is indifferent to challenging and not challenging with the high signal if

$$-L^H + \frac{\delta(\varepsilon)}{\psi(\varepsilon)} \frac{P_A + 2F}{P_A + F - P_{20}} R^{L|L} = \frac{\delta(\varepsilon)}{\psi(\varepsilon)} \tag{A.4}$$

where, as before, $\psi(\varepsilon) = \varepsilon^2 + (1-\varepsilon)^2$ is the probability that the signals are the same for a given ε and $\delta(\varepsilon) = 2\varepsilon(1-\varepsilon)$.

A.1.4. Seller's Indifference Between Challenging and Not Challenging After a High Signal. Panel (b) of Figure A.2 shows the expected value for challenging and not challenging for states of the world where the seller has a low signal and observes a high announcement. As before, the seller's likelihood of reaching each potential state depends on L^L whereas the expected value within these nodes depends on $R^{L|L}$. A seller is indifferent to lying and not lying if

$$-L^H + \frac{\psi(\varepsilon)}{\delta(\varepsilon)} \frac{P_A + 2F}{P_A + F - P_{20}} R^{L|L} = \frac{\psi(\varepsilon)}{\delta(\varepsilon)}. \tag{A.5}$$

Note that this is identical to the seller's indifference condition for challenging with the low signal except that the ratio of states is inverted.

Given the four indifference conditions, the point predictions of the model come from solving the four-by-four system of simultaneous equations. The solution to this system is as follows.

Result A.1. *With selfish agents, the mixed strategy equilibrium with* $\varepsilon = 0.05$ *is* $L^H = 0, R^{L|L} = 0.53333, C^{L|H} = 0.66,$ *and* $C^{L|L} = 0.285$. *The mixed strategy equilibrium with* $\varepsilon = 0.1$ *is* $L^H = 0$, $R^{L|L} = 0.53333$, $C^{L|H} = 0.625$, *and* $C^{L|L} = 0.625$.

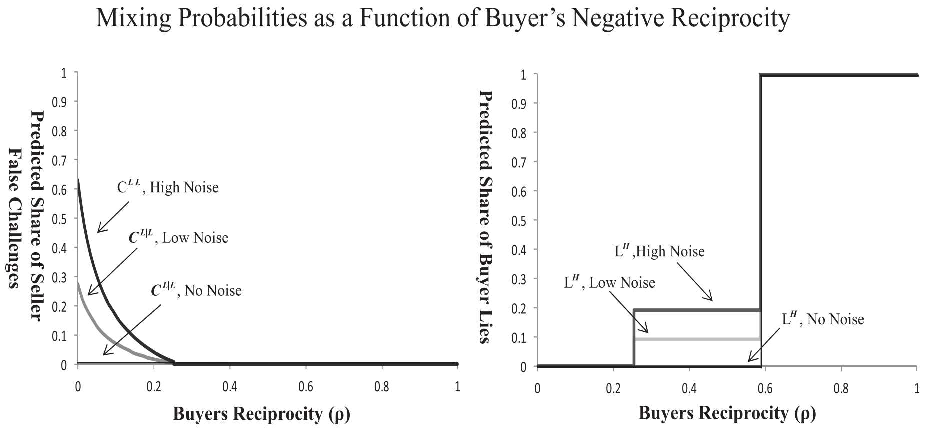
The surprising restriction that $L^H = 0$ is due to the fact that the seller must be indifferent to mixing in the case of a high and low signal.

In the next section, we show that when buyers have negative reciprocity and wish to retaliate against the seller for a challenge, the buyer may strictly prefer to reject after a low signal and a low announcement. This (likely) scenario eliminates seller false challenges (i.e., sets $C^{L|L}$ to zero) and instead leads to buyer challenges.

A.2. Point Predictions with Negative Reciprocity

In Fehr et al. (2014) (henceforth FPW), it is shown that buyers who view a seller's challenge as an unkind act may retaliate against the sellers by rejecting appropriate counter-offers where the expected value for accepting the counter-offer is small. Although the parametrization of the Moore-Repullo mechanism is set to make such retaliation unlikely without noise, there is no easy way to avoid retaliation from affecting the point predictions of the mixed strategy equilibrium where, by construction, the buyer is indifferent between accepting and rejecting a challenge. In this section, we discuss how the point predictions of the model changes when buyers become more reciprocal and the level of buyer reciprocity is common knowledge.

Following Dufwenberg and Kirchsteiger (2004), FPW shows that in a psychological games framework, a challenge by the seller is always seen by the buyer as an unkind act. Buyers who are prone to negative reciprocity may gain a "psychological" payoff by reducing the payoff of the seller and rejecting the counter-offer. Rather than



Mixing Probabilities as a Function of Buyer's Negative Reciprocity

FIGURE A.3. Point predictions for seller false challenges and buyer lies as a function of buyer reciprocity.

reconstructing the entire arguments of these previous works, we use this insight in a reduced form way. Let the utility of a buyer who rejects the seller's counter-offer be $-F + 75\rho$, where ρ is the additional utility the buyer receives from rejecting the sellers counter-offer and reducing the seller's payoff by 75. Note that since 75ρ is the amount of money that the buyer is willing to leave on the table to reject a counter-offer of the seller, ρ can be thought of as the amount $x that the buyer is willing to forego to punish the seller by $1.

Figure A.3 maps the point predictions of the four mixing probabilities as ρ increases.[35] As can be seen by looking at the point prediction of $C^{L|L}$, negative reciprocity initially reduces the proportion of false challenges. Such reductions in the likelihood of challenges increase the utility of the buyer for accepting the counter-offer and offset the utility from retaliation. As ρ increases, there exists a cutoff point for which the buyer will reject the counter-offer even if the seller never lies. At this point $C^{L|L} = 0$ and mixing occurs only over L^H and $C^{L|H}$. For very high ρ, outside the range of parameters seen in FPW, negative reciprocity can lead to the buyer rejecting challenges even when he has the high signal.

Based on the levels of negative reciprocity estimated using data from FPW, we expect ρ to range between 0.17 and 0.46. At these parameter values, mixing typically occurs over L^H and $C^{L|H}$. Note, however, that for all reciprocity levels below $\rho = 0.6$, the overall amount of buyer lies (L^H) plus seller false challenges ($C^{L|L}$) is unambiguously increasing in noise.

[35] Graphs are shown for the case where sellers are not reciprocal. In the more general case, seller reciprocity will reduce the frequency of buyer lies in cases where they arise in equilibrium. However, it has no effect on the ordering of treatments.

A.3. Pure Strategy Equilibrium

In the noise treatment, our game also admits two pure-strategy equilibrium in which buyers always lie. First, one can sustain the following "bad" (sequential) equilibrium with the appropriate sequence of beliefs: B announces low (i.e., a value of 20 ECU) in stage 1 regardless of his signal, S never challenges in stage 2, and (off-equilibrium) B always rejects a counter-offer made in stage 3 if that stage were to be reached.

More specifically, this equilibrium can be sustained as a sequential equilibrium with the buyer's (off-equilibrium) belief that the true state is low ($\theta = \theta^L$) when he is challenged and the arbitrator's counter-offer is made. To establish sequential rationality, we proceed by backward induction. It stage 3, regardless of his signal, B believes with probability one that the state is θ^L. Accepting S's offer at a price of 25 (resp. 75) leads to a payoff of $20 - 25 - 25 = -30$ (resp. $20 - 25 - 75 = -80$) whereas rejecting it leads to a payoff of -25. Thus, it is optimal for B to reject the offer. Moving back to stage 2, if S chooses "Challenge", S anticipates that her offer will be rejected by B in stage 3, and thus anticipates that, as ε goes to zero, the payoff is approximately equal to -25 if her signal is high and to -25 if the signal is low. On the contrary, if S chooses "No Challenge", S guarantees a payoff of 10. Thus, regardless of her signal, it is optimal for S not to challenge. Moving back to stage 1, suppose first that B receives the high signal s_B^H. Then, as ε becomes small, B believes with high probability that the true state is θ^H so that his expected payoff from announcing "low" is close to $70 - 10 = 60$, greater than 35, which B obtains when announcing "high." Thus, it is optimal for B to announce "low". A similar reasoning applies if B receives the low signal s_B^L. Finally, consistency of beliefs follows by identical arguments to those in AFHKT (footnote 13). Thus, the above is indeed a sequential equilibrium.

Note that if some subjects play this equilibrium we should observe an increase in buyers' lies because they announce a low value after a high signal. If sellers are also coordinating on this equilibrium, we should never see challenges by the seller.

A second pure strategy (sequential) equilibrium can be sustained where the buyer always announces high regardless of his signal. In this equilibrium, the buyer's (off-equilibrium) belief is that the true state is high with probability 0.1 in stage 2 when he receives the low signal, announces a low valuation, and is challenged. The expected value for accepting the challenge is $0.9 \times -5 + 0.1 \times 45 = 0$. Thus, he is indifferent between accepting and rejecting the challenge. If in stage 1 the buyer believes that the seller will always challenge, the expected value of this sequence of play is -25. The buyer can do strictly better by announcing a high value with the low signal and thereby guarantee himself a return of $0.9 \times -15 + 0.1 \times 35 = -10$. Note that if buyers play this equilibrium we should see an increase in the proportion of buyers making high announcements with the low signal. Buyers taking this action should believe that they will be challenged if they make a low announcement.

Appendix B: Mechanism used to elicit incentive compatible beliefs

In the follow-up treatment with incentive compatible beliefs, we use the following belief elicitation game based on a mechanism developed by Savage (1971). For each potential combination of announcement and signal, buyers are asked to submit a belief, b, between 0 and 100 corresponding to the percentage chance that the seller will call in the arbitrator. A random number $c \in [0, 100]$ is then drawn by the computer that corresponds to the "computer's percentage chance of calling in the arbitrator."

At the end of the experiment, one of the periods is randomly selected for payment. Using an eight-sided dice, the main experiment is paid 50% of the time whereas each of the four potential beliefs are paid 12.5% of the time. If a belief elicitation game is selected, the belief elicitation game is resolved as follows. If $b \leq c$ the buyer is matched with the seller and his outcome is based on the arbitration decision of the seller. If the seller does not call the arbitrator, the buyer receives \$20. If, however, the seller calls the arbitrator, the buyer receives \$0. If $b > c$, the buyer is matched to the computer. The computer calls the arbitrator with probability $c/100$ and thus the buyer receives \$20 with probability $1 - (c/100)$ and \$0 otherwise.

The mechanism is similar to the Becker, DeGroot, and Marshack (1964) mechanism and is shown by Karni (2009) to induce truthful reporting of beliefs for rational agents with any von Neumann–Morgenstern utility function. Further, as individuals are paid only for the main experiment or the bonus game, there is no concerns about hedging. The mechanism and payment scheme are thus robust to heterogeneity in risk aversion and are incentive compatible.

As the belief elicitation mechanism is relatively complex, we provide extensive training with the mechanism before the start of the experiment. Buyers receive both written and oral instructions about the mechanism, which include a series of examples that make clear that under reporting or over reporting beliefs can lead to worse outcomes. Subjects are also told explicitly that it is best to write down their true belief. Following the instructions, subjects are also given a series of quiz questions about the elicitation mechanism where they must calculate various potential outcomes for truthfully reported and misreported beliefs.

References

Aghion, Philippe, Drew Fudenberg, Richard Holden, Takashi Kunimoto, and Olivier Tercieux (2012). "Subgame-Perfect Implementation Under Value Perturbations." *Quarterly Journal of Economics*, 127, 1843–1881.

Aghion, Philippe and Richard Holden (2011). "Incomplete Contracts and the Theory of the Firm: What Have We Learned Over the Past 25 Years?" *Journal of Economic Perspectives*, 25(2), 181–197.

Allen, Franklin (1987). "Discovering Personal Probabilities When Utility Functions are Unknown." *Management Science*, 33, 542–544.

Andreoni, James and Hal Varian (1999). "Pre-play Contracting in the Prisoners' Dilemma." *Proceedings of the National Acadamy of Science of the United States of America*, 96, 10933–10938.

Angrisani, Marco, Guarino Antonio, Steffen Huck, and Nathan Larson (2011). "No-Trade in the Laboratory." *BE Journal of Theoretical Economics (Advances)*, 11, 1–58.

Arifovic, Jasmina and John Ledyard (2004). "Scaling up Learning Models in Public Good Games." *Journal of Public Economic Theory*, 6, 203–238.

Attiyeh, Greg, Robert Franciosi, and R. Mark Isaac (2000). "Experiments with the Pivot Process for Providing Public Goods." *Public Choice*, 102, 95–114.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak (1964). "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science*, 9, 226–232.

Blanco, Mariana, Dirk Engelmann, Alexander K. Koch, and Hans-Theo Normann (2010). "Belief Elicitation in Experiments: Is there a Hedging Problem?" *Experimental Economics*, 25, 412–438.

Bracht, Juergen, Charles Figuis, and Marisa Ratto (2008). "Relative Performance of Two Simple Incentive Mechanisms in a Public Goods Experiment." *Journal of Public Economics*, 92, 54–90.

Burford, Ingrid and Tom Wilkening (2016). "Belief Elicitation using the Stochastic Becker-DeGroot-Marschak Mechanism: A Comparison of Three Elicitation Proceedures." Working paper.

Carrillo, Juan D. and Thomas R. Palfrey (2011). "No Trade." *Games and Economic Behavior*, 71, 66–87.

Chen, Yan and Robert Gazzale (2004). "When Does Learning in Games Generate Convergence to Nash Equilibria? The Role of Supermodularity in an Experimental Setting." *American Economic Review*, 94(5), 1505–1535.

Chen, Yan and Charles Plott (1996). "The Groves–Ledyard Mechanism: An Experimental Study of Institutional Design." *Journal of Public Economics*, 59, 335–364.

Chen, Yan and Fang-Fang Tang (1998). "Learning and Incentive-Compatible Mechanisms for Public Goods Provision: An Experimental Study." *Journal of Political Economics*, 106, 633–662.

Cheung, Stephen L., Morten Hedegaard, and Stefan Palan (2014). "To See is to Believe: Common Expectations in Experimental Asset Markets." *European Economic Review*, 66, 84–96.

Chung, Kim Sau and Jeffrey Ely (2003). "Implementation with Near-Complete Information." *Econometrica*, 71, 857–871.

Datta, Somnath and Glen Satten (2005). "Rank-Sum Tests for Clustered Data." *Journal of the American Statistical Association*, 471, 908–915.

de, Clippel Geoffroy, Kfir Eliaz, and Brian Knight (2014). "On the Selection of Arbitrators." *American Economic Review*, 104, 3434–3458.

DuCharme, Wesley M. and Michael L. Donnell (1973). "Intrasubject Comparison of Four Response Modes for "Subjective Probability"Assessment." *Organizational Behavior and Human Performance*, 10, 108–117.

Dufwenberg, Martin and Georg Kirchsteiger (2004). "A Theory of Sequential Rationality." *Games and Economic Behavior*, 47, 268–298.

Ederer, Florian and Ernst Fehr (2009). "Deception and Incentives: How Dishonesty Undermines Effort Provision." Working Paper No. 341. IZA Discussion Paper No. 3200.

Falkinger, Josef, Ernst Fehr, Simon Gächter, and Rudolf Winter-Ebrner (2000). "A Simple Mechanism for the Efficient Provision of Public Goods: Eexperimental Evidence." *American Economic Review*, 90(1), 247–264.

Fehr, Ernst, Michael Powell, and Tom Wilkening (2014). "Handing Out Guns at a Knife Fight: Behavioral Limitations of Subgame Perfect Implementation." CESIFO Working paper No. 4948, CESIFO Group Munich.

Fischbacher, Urs (2007). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10, 171–178.

Fudenberg, Drew, David M. Kreps, and David K. Levine (1988). "On the Robustness of Equilibrium Refinements." *Journal of Economic Theory*, 44, 354–380.

Fudenberg, Drew and David K. Levine (1993). "Self-Confirming Equilibrium." *Econometrica*, 61, 523–545.

Gneezy, Uri (2002). "Deception: The Role of Consequences." *American Economic Review*, 95(1), 384–394.

Greiner, Ben (2004). “The Online Recruitment System ORSEE 2.0—A Guide for the Organization of Experiments in Economics.” *Working Paper Series in Economics 10*, University of Cologne, Department of Economics.

Grether, David M. (1981). “Financial Incentive Effects and Individual Decision Making.” *Social Science Working Paper 401*, California Institute of Technology.

Grossman, Sanford J. and Oliver Hart (1986). “A Theory of Vertical and Lateral Integration.” *Journal of Political Economy*, 94, 691–719.

Harstad, Ronald M. and Michael Marrese (1981). “Implementation of Mechanism by Processes: Public Good Allocation Experiments.” *Journal of Economic Behavior & Organization*, 2, 129–151.

Harstad, Ronald M. and Michael Marrese (1982). “Behavioral Explanations of Efficient Public Good Allocations.” *Journal of Public Economics*, 19, 367–383.

Hart, Oliver and John Moore (2003). “Some (Crude) Foundations for Incomplete Contracts.” Working paper.

Healy, Paul J. (2006). “Learning Dynamics for Mechanism Design: An Eexperimental Comparison of Public Goods Mechanisms.” *Journal of Economic Theory*, 129, 114–149.

Holt, Charles A. (2006). *Markets, Games, & Behavior*. Pearson/Addison Wesley, Boston.

Holt, Charles A. and Angela M. Smith (2016). “Belief Elicitation with a Synchronized Lottery Choice Menu That is Invariant to Risk Attitudes.” *American Economic Journal: Microeconomics*, 8, 110–139.

Hoppe, Eva I. and Patrick W. Schmitz (2011). “Can Contracts Solve the Hold-Up Problem? Experimental Evidence.” *Games and Economic Behavior*, 73, 186–199.

Huck, Steffen and Georg Weizsäcker (2002). “Do Players Correctly Estimate What Others Do?: Evidence of Conservatism in Beliefs.” *Journal of Economic Behavior & Organization*, 47, 71–85.

Kalai, Ehud and Ehud Lehrer (1993). “Rational Learning Leads to Nash Equilibrium.” *Econometrica*, 61, 1019–1045.

Karni, Edi (2009). “A Mechanism for Eliciting Probabilities.” *Econometrica*, 77, 603–606.

Kartik, Navin, Olivier Tercieux, and Richard Holden (2014). “Simple Mechanisms and Preferences for Honesty.” *Games and Economic Behavior*, 83, 284–290.

Katok, Elena, Martin Sefton, and Abdullah Yavas (2002). “Implementation by Iterative Dominance and Backward Induction: An Experimental Comparison.” *Journal of Economic Theory*, 104, 89–103.

Lei, Vivian, Charles N. Noussair, and Charles R. Plott (2011). “Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality.” *Econometrica*, 69, 1327–1347.

Maskin, Eric (1977. Published 1999). “Nash Equilibrium and Welfare Optimality.” *Review of Economic Studies*, 66, 39–56.

Maskin, Eric and Jean Tirole (1999a). “Two Remarks on the Property-Rights Literature.” *Review of Economic Studies*, 66, 139–49.

Maskin, Eric and Jean Tirole (1999b). “Unforseen Contingencies and Incomplete Contracts.” *Review of Economic Studies*, 66, 39–56.

Masuda, Takehito, Yoshitaka Okano, and Tatsuyoshi Saijo (2014). “The Minimum Approval Mechanism Implements the Efficient Public Good Allocation Theoretically and Experimentally.” *Games and Economic Behavior*, 83, 73–85.

Miettinen, Topi (2013). “Promises and Conventions—An Approach to Pre-Play Agreements.” *Games and Economic Behavior*, 80, 68–84.

Moore, John (1992). *Advances in Economic Theory: Sixth World Congress, Vol. I.* Cambridge University Press, pp. 182–282.

Moore, John and Raphael Repullo (1988). “Subgame Perfect Implementation.” *Econometrica*, 56, 1191–1220.

Nöldeke, Georg and Klaus Schmidt (1995). “Option Contracts and Renegotiation: A Solution to the Hold-Up Problem.” *RAND Journal of Economics*, 26, 163–179.

Ponti, Giovanni, Anita Gantner, Dunia López-Pintado, and Robert Mongtgomery (2003). “Solomon’s Dilemma: An Experimental Study on Dynamic Implementation.” *Review of Economic Design*, 8, 217–239.

Sanchez-Pages, Santiago and Marc Vorsatz (2007). “An Experimental Study of Truth-Telling in a Sender-Receiver Game.” *Games and Economic Behavior*, 61, 86–112.

Savage, Leonard J. (1971). “Elicitation of Personal Probabilities and Expectations.” *Journal of the American Statistical Association*, 66, 783–801.

Sefton, Martin and Abdullah Yavas (1996). “Abreu–Matsushima Mechanisms: Experimental Evidence.” *Games and Economic Behavior*, 16, 280–302.

Selten, Reinhart (1975). “Re-Examination of the Perfectness Concept for Equilibrium Points in Extensive Games.” *International Journal of Game Theory*, 4, 25–55.

van, Damme Eric (1984). “A Relation between Perfect Equilibria in Extensive form Equilibria in Normal Form Games.” *International Journal of Game Theory*, 13, 1–13.

Supplementary Data

Supplementary data are available at *JEEA* online.