

Agent Selection and Belief Polarization in Distributive Bargaining

David Hagmann¹ and Daniel C. Feiler²

¹The Hong Kong University of Science and Technology

²Tuck School of Business, Dartmouth College

Author Note

Correspondence concerning this article should be addressed to David Hagmann, The Hong Kong University of Science and Technology, Department of Management, The Hong Kong University of Science and Technology, Hong Kong, Email: hagmann@ust.hk

Abstract

Many negotiations—from legal disputes to labor contracts—are conducted through agents rather than directly by principals. While significant attention has been given to the misalignment of interests between principals and agents, little is known about how the process of selecting an agent affects bargaining outcomes. Across four preregistered experiments ($N = 4,385$), we show that principals systematically choose agents who go on to make overly aggressive offers. Principals on each side preferentially select agents whose beliefs about fair and achievable outcomes are especially favorable to their respective side. Thus, they send pairs of agents to the bargaining table whose expectations are strongly opposed and who are therefore more likely to reach impasse. Agent selection leads to worse outcomes than if principals chose randomly or negotiated on their own behalf, and outcomes could be improved by unilaterally selecting less aggressive agents. We document these patterns across distinct bargaining environments: an ultimatum-style settlement task and a real-time, open-ended negotiation.

Keywords: bargaining, principal-agent problem, confirmation bias, belief polarization

Agent Selection and Belief Polarization in Distributive Bargaining

Introduction

Bargaining is a common aspect of contracting, economic exchange, and dispute resolution, with a rich history in both theoretical and experimental research (Eisenhardt, 1989; Güth et al., 1982; Nash, 1950). Prior work has largely characterized and tested the behavior of people engaging in negotiations and ultimatum games on their own behalf. In practice, however, bargaining is frequently conducted through agents rather than by the principals themselves (Mnookin & Susskind, 1999; Schotter et al., 2000). The principal-agent relationship creates a well-studied challenge, which is that the interests and incentives of the agent may not be aligned with those of the principal whom he or she represents (Ross, 1973). In this paper, we examine the consequences of a different, important, and thus far unexplored feature of the principal-agent relationship: the process of choosing an agent from a set of possible agents.

Agent-selection precedes interactions conducted through intermediaries. In the simplest case, principals may use the first agent they encounter or the only agent available to them. However, in many cases, particularly those involving high stakes, principals may be motivated to seek out and interact with multiple agents prior to selecting one. These prospective agents will naturally vary in their beliefs about what a fair outcome would be, as well as in their claims to the principal of what bargaining outcome they can achieve if selected. Having courted the set of potential agents who vary in their beliefs and claims, principals must then select one of these agents to represent them. Often the principals do not have sufficient expertise or information to have full confidence in their independent assessment and draw inferences about the likely outcome based on the claims they are presented with by the potential agents.

We hypothesize that principals engaging in such an agent-selection process pick overly aggressive agents to their own detriment. We propose that principals tend to select an agent who holds beliefs about fair and achievable bargaining outcomes that are skewed in favor of the principals' position, and more so than the pool of possible agents in general. This tendency occurs for both parties, but in opposite directions given their opposing incentives. Consequently the

selected agents who ultimately meet at the bargaining table are particularly polarized in what they believe a fair and achievable resolution would be. With both opposing agents then believing that their walkaway option is more favorable than what the other party is offering, negotiations are more likely to end in impasse, imposing costly consequences on both parties. Across four multi-stage, pre-registered experiments ($N = 4,385$), we find evidence of this agent-selection dynamic and show how it can negatively affect bargaining outcomes by inflating impasse rates, both in a stylized ultimatum setting and in a live, open-ended negotiation.

Costly impasse has been documented across a wide-variety of managerial domains, including contractual negotiations, collective bargaining agreements, and malpractice disputes ([Ashenfelter & Currie, 1990](#); [Card, 1990](#); [Kennan & Wilson, 1990](#)). Some impasses lead to foregone gains, as mutually beneficial exchanges fail to materialize, and others lead to costly, ongoing disputes. In 2016, Plaintiffs and Defendants in tort claims paid \$135 billion in legal fees, amounting to approximately a third of the total cost of torts ([U.S. Chamber Institute for Legal Reform, 2018](#)). In this paper, we present evidence suggesting that the agent-selection process may contribute to the high incidence rate of such costly failures to reach agreement.

There are, of course, many reasons why working through agents is beneficial and therefore widespread in practice and widely studied across disciplines. For example, the literature has discussed at length how agents may have more expertise, information, or lower time costs. Our work suggests that the choice of a specific agent, however, may have gone unnoticed as an important driver of excessive impasse in bargaining practice.

The dynamics we study arise in a range of environments where principals actively select agents prior to negotiation. In legal disputes, plaintiffs and defendants each retain counsel from among multiple prospective attorneys who vary in their assessments of the case's likely outcome. In business-to-business contracting, firms hire intermediaries or brokers to negotiate terms in procurement. In labor negotiations, both unions and management select representatives to bargain on their behalf. In mergers and acquisitions, parties engage investment bankers or advisors whose projections of deal value may influence both selection and subsequent bargaining posture.

Consider, for instance, a plaintiff evaluating prospective attorneys for a personal injury claim: one attorney projects a settlement of \$200,000 while another projects \$250,000. The principal, lacking independent expertise, may understandably gravitate toward the more optimistic projection. Indeed, the principal could convince themselves of many possible reasons why the more optimistic attorney is the better choice—they may be more informed, more motivated, have previously secured larger settlements, etc.—although such reasoning may be motivated by the fact that the more optimistic attorney is giving them “good news.” Importantly, the defendant in the dispute will go through a similar process in the opposite direction, looking for an attorney who can minimize the cost of the settlement. When both sides engage in this process, the attorneys who ultimately meet at the bargaining table are not random draws of all attorneys, whose average judgment may be accurate, but agents who went through a selection process based on their beliefs. The selected agents thus hold more polarized expectations than the broader pool of available counsel, increasing the likelihood of costly impasse. This polarization through selection occurs potentially in addition to strategic shading that agents may engage in, where they might promise their clients more favorable outcomes in an effort to be selected. Our experiments are designed to isolate and test this structural feature common across these settings: principals select agents based on projected outcomes, and that selection process shapes which beliefs and strategies are brought to the bargaining table.

Bargaining, Agency, and Ultimatums

The behavioral and strategic dynamics of bargaining have long been of interest to management science (Gächter & Riedl, 2005; Harsanyi & Selten, 1972; Nagarajan & Bassok, 2005; Rangaswamy & Shell, 1997; Samuelson, 1980; Shachat & Swarthout, 2013; Shachat & Tan, 2015; Tang et al., 2009). More recent work has focused on how bargaining affects supply chain coordination and efficiency (Davis & Hyndman, 2019, 2021; Haruvy et al., 2020; Leider & Lovejoy, 2016). The theoretical literature on principal-agent relationships has emphasized the challenge of aligning interests through incentives (Grossman & Hart, 1983; Ross, 1973). Meanwhile, studies in the laboratory have tested the social and psychological effects of agency

itself. For example, previous work has found that when representing the interests of another person (as opposed to representing their own interests), women anticipate and face less social “backlash” (i.e., dislike) for exhibiting assertiveness ([Amanatullah & Morris, 2010](#); [Amanatullah & Tinsley, 2013](#)). In the setting of a dictator game, research has found that agents implement more selfish decisions on behalf of the principal than do the principals themselves ([Hamman et al., 2010](#)). Principals benefit further because third party observers tend to punish the agent who implemented the selfish choice rather than the principal who benefited from it ([Bartling & Fischbacher, 2012](#)). Further, [Coffman \(2011\)](#) found that selfish actions are punished less—and generous actions are rewarded less—if implemented through an intermediary rather than by the principal directly. Even when the principal acts on their own behalf, merely receiving nonbinding advice to implement a selfish option reduces third party perception of their selfishness and culpability ([Coffman & Gotthard-Real, 2019](#)).

In past research on the ultimatum game, agents acting on behalf of ultimatum senders made more generous offers than the senders themselves, and ultimatum receivers were more likely to accept them ([Fershtman & Gneezy, 2001](#)). When agents competed to be hired as representatives in an ultimatum game, they strategically communicated lower offers than they implemented after being hired ([Choy et al., 2016](#)). The same study, notably, also found that the hired agents’ offers were more generous than what the principals implemented on their own. Contrary to our hypothesis, this might suggest that agents reduce impasse rates. In the context of a distributive bargaining game, however, agents negotiating on behalf of principals have been found to be less likely to reach agreement than principals negotiating with each other directly ([Schotter et al., 2000](#)). This finding held irrespective of whether agents received a share of the realized surplus or a fixed payment for reaching an agreement. Moreover, merely placing experimental participants into the role of agents does not make them less biased than if they played the role of principals ([Melnikoff & Strohminger, 2020](#)), suggesting that agents may be just as susceptible as principals to self-serving interpretations of information and fairness ([Babcock et al., 1995](#)). Indeed, delegating a task to an agent may increase bias (see [Burson et al., 2010](#) for such a

demonstration outside of the context of bargaining). However, these comparisons assume an existing principal-agent relationship (or, in the experimental context, randomly matched participants) and do not account for the selection of agents and its potential consequences for competitive interactions, a gap we address in this paper.

When agents compete to be hired, they may attempt to provide information that will be most appealing to the principal. Hamman et al. (2010) show that people prefer to hire agents who indicate that they will implement a selfish decision in a dictator game. In an effort to be hired, agents may simply present information that confirms the principal's priors, thereby delivering the news they most wish to receive (see Cummins & Nyman, 2013 for a theoretical approach). Evidence from the laboratory indeed suggests that some principals prefer to hire agents who withhold information about the negative externalities of the choices the principals make (Shalvi et al., 2019). Reliance on an agent may thus serve as a strategy to avoid information that the principal would prefer not to have (Golman et al., 2017). We propose that this desire for good news may also affect the selection of agents based on what they project they can achieve on the principal's behalf.

Taken together, previous work illustrates how behavior and the perceptions of behavior shift as a function of whether a principal acts on her own behalf or whether she is represented by an agent. However, absent from this discussion is the fact that agents in the real world are endogenously selected from a pool of contenders and little is known about how principals take into account agents' beliefs in their selection and how such agent selection affects bargaining. This choice, of course, can substantially shape the outcome of any bargaining process. Moreover, choices by principals in the aggregate also determine which agents remain in the market and therefore are available for subsequent selection. Mistakes people make when selecting agents to act on their behalf may therefore not only harm their own bargaining outcomes but also adversely affect future principals.

Selecting Polarized Agents and its Consequences

We present results from four large, preregistered experiments ($N = 4,385$). In Studies 1–3, participants read materials from a real personal injury lawsuit (Babcock et al., 1995) and engage in an ultimatum-style settlement game; in Study 4, they engage in live, open-ended negotiation over a commercial transaction between two businesses. In Study 1, we find that the agents who get selected hold more polarized beliefs than agents overall. This increases the rate of impasse, making both parties worse off than if they selected an agent at random, rather than engaging in this deliberative (but biased) agent-selection process. However, selecting a random agent is not an equilibrium. We characterize a game using the empirical payoffs and find that the strategic incentives mirror that of a prisoner’s dilemma: while both parties would be better off choosing an agent at random (“cooperating”), the sole Nash equilibrium is to engage in deliberate agent-selection (“defecting”). This result is consistent with evidence from the field, where retaining lawyers often leads to worse outcomes for both parties than not retaining lawyers—but when only one party uses an agent, it benefits them at the other’s expense (Ashenfelter et al., 2013). In the real world, we cannot observe the counterfactual outcome of choosing a less aggressive agent. In the context of our experiment, however, we can see that principals could improve their payoff by unilaterally choosing a less aggressive agent. Thus, the costs arise not merely from agent-selection itself, but from selecting agents with overly favorable beliefs about the outside option obtained in case of an impasse.

One explanation for the overly aggressive choice of agent is that the principals themselves are more biased than the average agent (e.g., because they are emotionally invested in the outcome). Thus, they may merely select an agent who matches their own beliefs. In that case, agent selection would fail to attenuate principals’ bias, but not itself worsen outcomes relative to self-representation. However, in Study 2, we show that principals select representatives with beliefs about what they can achieve, as well as beliefs about the outside option in case of an impasse, that are more favorable than what the principals themselves believe. Since this selection of overly aggressive agents harms performance, in our study bargaining through selected agents

yields worse outcomes than self-representation.

One reason principals in the real world retain agents is because of information asymmetries: the agents may be more informed and knowledgeable than the principals. Moreover, agents succeed and persist not at random, but need to be retained in order to remain in business and hence available to principals. In Study 3, we extend our findings into a setting with information asymmetry between principals and agents, where the latter is more informed. We show that engaging in agent-selection makes principals themselves more polarized in their fairness beliefs and expectations about the outcome of the case. Moreover, we show how the pool of available agents changes over time when prior hiring choices affect future availability. Specifically, in our study agents who are selected “survive” and remain available in the next period (following [Axelrod & Hamilton, 1981](#)). The least aggressive agents tend not to be selected and disappear from the market, which leads to an increasingly polarized pool of potential agents. Principals in later rounds neglect to account for this selection, even when they are fully informed about the process, and continue to choose relatively aggressive agents from their available choice set. Therefore, the rate of impasse from excessive polarization of selected agents increases over time.

Finally, in Study 4, we extend our findings to a substantively different bargaining environment to examine the robustness of the observed agent-selection dynamics. While the previous studies rely on a legal context and an ultimatum game structure, participants in this experiment engage in a live, text-based, open-ended negotiation over the sale of a piece of technology. Both parties can make, counter, accept, and reject offers over multiple rounds of communication. This provides an opportunity for aggressive agents to make strong initial offers to anchor their counterpart, then moderate their demands and reach agreement before the negotiation ends. Contrary to this possibility, we find that agent selection continues to increase impasse rates in this more naturalistic setting.

Open Science Statement

We report all hypotheses, data exclusions, manipulations, and measures in the studies. Screen captures of the experimental materials are available in the Supplementary Materials. The preregistration reports, the complete data, and the code to reproduce all statistical analyses and figures in the manuscript are available via OSF.¹ The negotiation platform designed for Study 4 is available via Github.²

Study 1: Selected Agents vs. All Potential Agents

We randomly assign participants to one of two parties in a dispute (Plaintiff or Defendant) and one of two roles (Principal or Agent) in a 2×2 design. All parties read identical case materials and learn that they will engage in an ultimatum settlement negotiation. Agents send non-binding signals about the outcome they think they can achieve if they are hired, and Principals choose an Agent on the basis of this signal. Prospective Agents, prior to learning whether they had been selected, submit an ultimatum offer and rejection threshold for a settlement offer. Only selected Agents are matched with each other and if the offer is rejected, the decision of a neutral third party (a real-world retired judge) is implemented, along with a symmetric penalty for both parties representing the cost of delay and litigation.

A key innovation of our design is that the ultimatum negotiation allows us to simulate bargaining outcomes for any pair of Agents in the experiment. We can then compare the realized outcomes, where both parties engage in agent selection, against the counterfactual of random matching. Further, we calculate whether, conditional on agent selection, principals are choosing the ex-post optimal agent. Although we do not expect participants to intuit who this agent would be, it serves as an interesting and generally informative benchmark for agent selection. We predict that principals on both sides systematically deviate from this ex-post optimal choice by choosing an agent who holds a belief about the fair outcome that is more favorable to their position than the agent who would have maximized their expected earnings. That is, we expect that selecting

¹ https://osf.io/s6kaz/?view_only=b98edbd7c87a4029b5c5ee8a38163602.

² <https://github.com/dhagmann/negotiation-platform>.

Agents with overly favorable outcome projections leads to impasse and is ultimately costly for both parties.

Experimental Design

We invited participants from Amazon Mechanical Turk in two stages. In the first stage, we recruited participants for the role of Agents (lawyers) in a personal injury lawsuit and randomly assigned them to represent the Plaintiff or the Defendant. In the second stage, we recruited new participants for the role of Principals and randomly assigned them to be Plaintiffs or Defendants. Principals observed decisions made by Agents and we therefore conducted these two stages sequentially. The study took place on consecutive days. Each participant received a fixed payment of \$1, a fixed incentive of \$1 for passing a comprehension check and advancing to the main part of the study, plus any additional earnings from the experiment. The median completion time was approximately 30 minutes and participants who passed the comprehension check and completed the full study earned an average of \$4.69. We preregistered our hypotheses, sample size, exclusion criteria, and analyses on AsPredicted.³

Agents

Participants in the first stage were randomly assigned to the side of the Plaintiff or the Defendant and learned that they were competing with four other participants to be hired as an attorney. If they were hired, they would receive an additional \$1 bonus and would be matched with an Agent for the opposing party who was selected via the same process, making them eligible for additional earnings. The hiring decision would be made by new participants in the role of either Plaintiff or Defendant. Prospective Agents were informed that they could later send one (non-binding) signal to the Principals: the outcome of a distributive bargaining game they believed they could achieve on the principal's behalf.

Participants then read materials related to a personal injury lawsuit resulting from a traffic collision, taken from Babcock et al. (1995).⁴ The case materials consisted of 10 documents

³ https://aspredicted.org/blind.php?x=PAQ_LDR

⁴ We are grateful to the authors for sharing the case materials with us.

containing a rich set of information: the Plaintiff’s initial claim, the Defendant’s response, both of their court testimonies, the testimonies of experts and witnesses, and their respective driving records. After reading all materials, participants answered five comprehension check questions about the content, which were identical across the roles. Participants who did not answer all five questions correctly on either the first or second attempt were excluded from the study and we did not collect further information from them.

Next, participants sent to the Principal the outcome they believed they could achieve if they were hired as their Agent (which we will refer to as *Agents’ signals*). They submitted an amount between \$0 and \$100,000. They were reminded that this amount would be communicated to a Principal who would choose one of five Agents based only on this information. After submitting the signal, without forewarning, they received an additional bonus opportunity to estimate the ruling of a real, neutral judge who rendered a decision based on the same experimental materials (*beliefs about fair outcome*). Participants could earn an additional 50 cents if they were within \$5,000 of what the judge had decided would be a fair outcome.

Finally, we introduced participants to the ultimatum negotiation game. They were reminded that, if they were hired by a Principal, they would be matched with another participant who had been hired as an Agent for the opposing side. Either they or the other Agent would then get to make a settlement offer to the other party, who got to decide whether to accept or reject the offer. If the offer is accepted, the case is considered “settled” and the earnings for both parties will be a function of the offered settlement. If the offer is rejected, the case would go to the judge who’s ruling they had just estimated. The judge’s ruling would determine the outcome, but both parties would further incur a 30% penalty to represent the cost of litigation. The payoffs for the two Agents were as follows:

$$\pi_{\text{Settlement}} = \begin{cases} \frac{\text{Settlement Amount}}{20,000}, & \text{for Plaintiff} \\ 4 - \frac{\text{Settlement Amount}}{20,000}, & \text{for Defendant} \end{cases}$$

$$\pi_{\text{Impasse}} = \begin{cases} \frac{\text{Judge's Ruling}}{20,000} \times 0.7, & \text{for Plaintiff} \\ \left(4 - \frac{\text{Judge's Ruling}}{20,000}\right) \times 0.7, & \text{for Defendant} \end{cases}$$

We set the initial endowment for the Defendant such that they could offer to “pay” to Plaintiffs most outcomes in the available range without incurring negative earnings. Only later would it be randomly determined which party would be the ultimatum giver and which the ultimatum receiver. Therefore, participants had to make two *bargaining decisions*: (1) They proposed a settlement amount in case they were selected to make the ultimatum offer, and (2) they selected a rejection threshold representing the lowest (Plaintiff) or highest (Defendant) amount they would accept in a settlement offer from the other side. The judge’s decision was \$30,560 (as reported in [Babcock et al., 1995](#)), which was not revealed to participants until the very end of the study. The experiment then concluded with basic demographic questions (gender, age, ethnicity, and education).

Principals

Participants in the roles of Principals were randomly assigned to the side of Plaintiff or Defendant. They were informed about their role and that they would get to hire an Agent from among five previous participants who had read the identical case materials they will be reading and passed the same comprehension check on these instructions as they will see. Moreover, participants were informed about the prospective Agents’ incentives, that is a fixed \$1 bonus from the experimenter for getting hired and identical earnings from the negotiation as the Principal who hires them. We then presented participants with the materials from the case and excluded those who did not answer the five-item comprehension check correctly on either the first or the second attempt.

Principals then received information about the nature of the ultimatum negotiation game, along with the signals sent by five participants from the previous stage. We ordered them from the lowest to the highest proposed outcome and referred to them as “Lawyer A” through “Lawyer E.” Note that for Plaintiffs, Lawyer E provided the estimate that was most favorable to their side (the

highest outcome), while for Defendants, the most favorable estimate came from Lawyer A (the lowest outcome). To ensure participants observed a representative sample of Agents, we partitioned the potential Agents into five quintiles on the basis of the signal they had sent to the Principal and selected one Agent at random from each quintile.⁵

After the principals had selected an Agent, there was a final bonus opportunity, without forewarning: they could earn a 50 cent bonus for estimating the judge's ruling within \$5,000. This provided us with an estimate of what they believe a "fair" outcome would be, after observing the signals from prospective Agents. The survey concluded with the same basic demographic questions we asked of the prospective Agents.

Matching

We randomly paired Selected Agents and further randomly picked one of them to be the ultimatum giver and the other to be the ultimatum taker. Selected Agents were paid based on the outcome and the Principals whom they represented received identical earnings. Because these pairings were done at random, we can reduce the noise in our measurement for the purpose of analysis by calculating the expected earnings for each Agent given all possible pairings, rather than the earnings from the realized pairing. All analyses are based on these expected earnings.

Results

We recruited 511 participants for the role of Agents. Of those, 39 did not pass the comprehension check and 50 provided an inconsistent offer in the ultimatum negotiation game.⁶ As preregistered, these participants were excluded, leaving us with 422 Agents. The following day, we recruited 504 new participants for the role of Principals. Of those, 30 failed the comprehension check, leaving us with a final sample size of 474 Principals.

⁵ We show the ranges covered by each quintile for the respective parties in the Appendix, Figure A1.

⁶ An inconsistent offer is one in which the rejection threshold is such that they would reject the offer they themselves have made. That is, they would have reached impasse when negotiating with themselves. The Agent for the Plaintiff, for example, might have offered a settlement of \$40,000, but said she would reject any offer less than \$50,000. We believe this is likely a result of confusion about the nature of the ultimatum settlement game.

Negotiation Outcomes

To estimate the effect of the agent-selection process on negotiation outcomes, we compare the average earnings for Plaintiffs and Defendants under two scenarios. First, we consider an outcome in which all possible Agents were selected exactly once. This is equivalent to a participant retaining an Agent at random or the real-world equivalent of retaining the first Agent she encounters. We pair every Agent for the Plaintiff against every Agent for the Defendant twice, once using their decision as the ultimatum giver and once using their decision as the ultimatum taker. We then average the earnings across all these pairs, providing us with the expected earnings for each Agent, conditional on neither party engaging in agent selection. Although the Principals in our experiment did not have the option to forgo the agent-selection process, this random pairing allows us to establish a benchmark for evaluating their performance with the deliberate choice they made in the experiment.

Second, we repeat the process above for the universe of *Selected Agents*. Agents were selected for display to Principals with replacement and therefore, an agent could have been selected by more than one Principal. Agents who were selected by multiple Principals are included the corresponding number of times, whereas Agents who were never selected by a Principal receive no weight in this analysis. This provides us with expected earnings conditional on both parties engaging in agent selection.⁷

We show the average earnings for all Agents as well as for selected Agents in Figure 1. We find that Principals' earnings are on average 5% lower with Selected Agents than if they were assigned an Agent at random ($t(448) = 6.74, p < .001$ for Plaintiffs and $t(400) = 5.38, p < .001$ for Defendants). The right panel of Figure 1 shows the impasse rate under each scenario. The impasse rate increases from 65% with All Agents to 79% with Selected Agents ($t(850) = 2.85, p = .005$). The agent-selection process leads to more impasse and makes both parties worse off

⁷ Consistent with our preregistered exclusion criterion for the analysis of all Agents, we exclude the 44 observations where Principals selected an Agent who exhibited misunderstanding of the ultimatum game in their ultimatum offer and rejection threshold. Since Principals did not observe the Agents' bargaining decisions, they could not have inferred this error. All our results are robust to including these inconsistent Agents and the Principals who chose them (see Table A2 in the Appendix for Study 1 and Tables B2 and B3 for Study 2, which used identical exclusion criteria).

Figure 1

Negotiation outcomes in Study 1. The first two panels show average earnings for Plaintiffs and Defendants, respectively. The third panel shows the impasse rate. Principals earn less when both parties engage in agent selection (orange) than if they chose at random (blue), driven by a higher rate of impasse. Error bars show 95% confidence intervals.

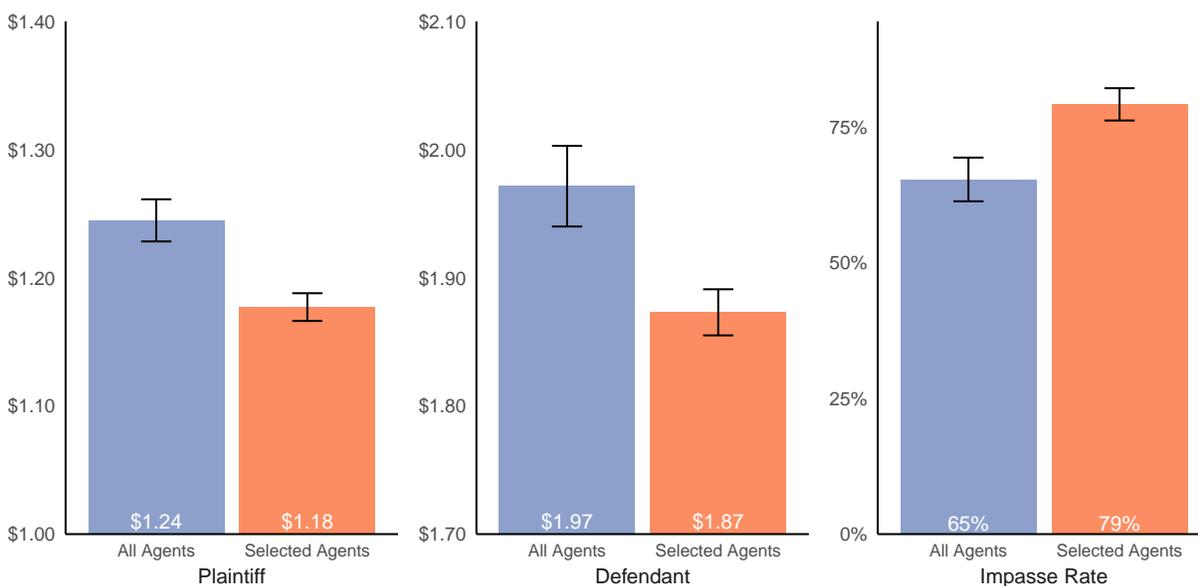


Figure 2

Selected Agents from each side hold more polarized beliefs than do Agents from each side overall in Study 1. Ex-post optimally selected Agents from each side hold beliefs that do not differ from one another. Circles show Agents for the Plaintiff, triangles show Agents for the Defendant. The dashed line shows the judge's decision and error bars show 95% confidence intervals.



than if had they retained an Agent at random.

Beliefs About Fair Outcome

When bargaining, beliefs about the value of the outside option are fundamental to whether an agreement is reached. Specifically, an Agent should turn down an offer if she believes it is less valuable than what she receives after impasse. That is, not every impasse is inherently a bad outcome for both parties. In our setting, the outside option is determined exogenously by the neutral judge's ruling. We next look at what may be driving the worse outcomes among the Selected Agents: more polarized beliefs about what the judge in the case would decide. Figure 2 shows the average private (and separately incentivized) estimate of the judge's ruling (the correct answer, \$30,560, is shown as a dotted line). Agents for the two parties differ in their estimate (\$32,652 for the Plaintiffs' Agents versus \$25,343 for those of the Defendants, $t(420) = 3.55$, $p < .001$), replicating prior work on the self-serving interpretation of information (Babcock et al., 1995).

Our key analysis relates to the pool of *Selected* Agents. We conduct an OLS regression with the role (Agent for the Plaintiff or Defendant) and the pool (All Agents or Selected Agents), as well as their interaction. We find that the Plaintiff-Defendant belief gap is significantly larger among the group of Agents who ultimately determine the outcome of the case, than among all potential Agents ($b = 8,375.30$, 95% CI [3,143.47, 13,607.12], $t(848) = 3.14$, $p = .002$). That is, the beliefs of Agents who were selected for bargaining to determine outcomes for the Principals are even more polarized than are the Agents overall.

Agents' Signals

When principals selected their Agent, they did so on the basis of a non-binding signal (the proposed outcome of the negotiation), without knowing how the Agent would act in the negotiation. This signal does not commit the Agent to any specific decision in the ultimatum game, which raises the question of whether Agents send informative signals, or whether they engage in cheap talk in the interest of being hired, only to act differently after being selected.

We find that the signals were indeed informative. Agents who signaled a higher outcome

to their prospective clients proposed higher amounts in the ultimatum offer ($r = .77$, 95% CI [.71, .82], $t(223) = 18.11$, $p < .001$ for Plaintiffs, and $r = .80$, 95% CI [.74, .84], $t(195) = 18.52$, $p < .001$ for Defendants), and set higher rejection thresholds ($r = .76$, 95% CI [.70, .81], $t(223) = 17.70$, $p < .001$ and $r = .71$, 95% CI [.63, .77], $t(195) = 13.98$, $p < .001$, for Plaintiffs and Defendants, respectively). Moreover, the signals correlated with their privately reported beliefs about the judge's ruling, which were incentivized for accuracy ($r = .66$, 95% CI [.58, .73], $t(223) = 13.12$, $p < .001$ and $r = .84$, 95% CI [.79, .88], $t(195) = 21.44$, $p < .001$ for Plaintiffs and Defendants, respectively). These results suggest that Principals are in fact selecting Agents who genuinely believe that they can achieve more favorable outcomes than others, causing them to subsequently make overly aggressive ultimatum offers and reject offers that would lead to better outcomes.

Ex-Post Optimal Strategy for Principals

Recall that Principals observed a representative set of five Agents from their respective side, with one Agent drawn from each of the signal quintiles (that is, each quintile of proposed outcomes—from lowest to highest). We now derive the empirical optimal quintile for the Principals, using the distribution of negotiation behavior by the Agents. Our design allows us to construct a payoff table with the expected earnings from selecting each quintile, given the quintile selected by the opposing Principal. We hypothesized systematic deviation from this rational expectations benchmark (Aumann, 1976; Muth, 1961), making a preregistered prediction that participants would choose from a quintile that is more favorable to their side than the ex-post optimal choice.

Because each side in the dispute can select from one of five quintiles, there are 25 possible quintile pairings that can meet at the negotiation table. Each of those pairings will have an expected payoff for the Plaintiff and the Defendant. Figure 3 shows the earnings for all the possible pairings. Along the x-axis, we vary which quintile the Plaintiff selects from, where a higher quintile corresponds to a higher signal, which is more favorable to the Plaintiff. Along the y-axis, we vary the quintile selected by the Defendant. A higher quintile similarly corresponds to

Figure 3

Expected payoffs for the Plaintiff and the Defendant when choosing an agent from each quintile, given each quintile the other party could have selected from in Study 1. The first number in each cell shows the earnings for the Defendant (row player), the second shows earnings for the Plaintiff (column player). Cell shading indicates the impasse rate (blue = low impasse, orange = high impasse). This empirical payoff table allows us to determine the unique pure strategy Nash Equilibrium: Plaintiffs should choose an agent from the third quintile and Defendants should choose an agent from the fourth quintile (marked in purple).

Defendant Quintile Selected	(5) Least favorable	(\$2.35, \$1.61) Impasse: 3%	(\$2.06, \$1.83) Impasse: 9%	(\$1.80, \$1.97) Impasse: 19%	(\$1.58, \$1.96) Impasse: 38%	(\$1.47, \$1.80) Impasse: 61%
	(4)	(\$2.71, \$1.14) Impasse: 13%	(\$2.32, \$1.30) Impasse: 32%	(\$2.03, \$1.34) Impasse: 53%	(\$1.82, \$1.27) Impasse: 76%	(\$1.74, \$1.17) Impasse: 91%
	(3)	(\$2.67, \$0.97) Impasse: 30%	(\$2.11, \$1.14) Impasse: 62%	(\$1.90, \$1.17) Impasse: 78%	(\$1.79, \$1.14) Impasse: 89%	(\$1.74, \$1.10) Impasse: 97%
	(2)	(\$2.51, \$0.93) Impasse: 46%	(\$1.98, \$1.09) Impasse: 78%	(\$1.83, \$1.11) Impasse: 88%	(\$1.75, \$1.11) Impasse: 94%	(\$1.72, \$1.11) Impasse: 97%
	(1) Most favorable	(\$2.14, \$1.00) Impasse: 72%	(\$1.82, \$1.11) Impasse: 89%	(\$1.76, \$1.12) Impasse: 93%	(\$1.73, \$1.12) Impasse: 96%	(\$1.72, \$1.11) Impasse: 97%
		(1) Least favorable	(2)	(3)	(4)	(5) Most favorable
		Plaintiff Quintile Selected				

a higher signal, but that is less favorable to the Defendant. Thus, the top-left cell shows that if the Defendant chooses the Agent from the least favorable quintile (5) and the Plaintiff chooses the Agent from the least favorable quintile (1), then the Defendant earns on average \$2.35 and the Plaintiff earns on average \$1.61, and only 3% of pairings end in impasse. Solving this game results in a unique, pure-strategy Nash Equilibrium, where the Defendant selects the 4th quintile (a less favorable Agent than the median) and the Plaintiff selects the 3rd quintile (the median Agent), shown in purple. This result provides a benchmark of “optimality” for which quintile of Agent the Principal should select.

We do not expect participants to have perfectly anticipated the decisions of others, thereby deciphering these exact equilibrium dynamics. Nor do we expect that this would be the ideal choice of quintiles in all settings.⁸ Rather, we compare the observed behavior to this benchmark because it provides useful insight with respect to how participant decisions are affecting bargaining outcomes. To the extent that Principals deviate from the optimal choice, they might do so randomly and therefore adding noise around the optimal choice. However, we predicted that they would systematically deviate from this benchmark, choosing an Agent from a quintile that is effectively more prejudiced to their side. We use this benchmark instead of a simpler alternative (e.g., “choosing the middle option”) because it is ex ante not clear what the correct behavior would be. For example, on an estimation task with no room for motivated reasoning, the median Agent would likely perform best (Neto & Fontanari, 2023). But since Agents interpret information self-servingly, it is not clear that this quintile would have the most accurate beliefs. Moreover, it is not clear ex ante that Agents with accurate beliefs would achieve the best outcome in a negotiation.

A key parameter influencing this equilibrium is the cost of litigation, which we fixed at 30% in this experiment. To assess robustness to differences to the costs of litigation, we recompute the payoff matrix and agent earnings under counterfactual rates of value destruction ranging from 0% (no cost) to 50%, holding agents’ settlement offers and acceptance thresholds fixed (Table 1).

⁸ Notably, however, the equilibrium remains unchanged in Study 2, which relies on the same experimental materials. This suggests that the equilibrium pairing is not purely a result of chance and may be a result of the incentives and the judge’s ruling to which the Agents respond in their decisions.

Table 1

Robustness of the ex post optimal pairing to varying levels of value destruction in Study 1. The 30% row (bold) reflects the actual experimental parameterization. t-tests compare per-agent mean earnings under random pairing (All) vs. observed participant selection (Selected).

Destruction	Nash Eq.	Plaintiff			Defendant		
		All	Selected	t-test	All	Selected	t-test
0%	(mixed only)	\$1.54	\$1.54	.800	\$2.46	\$2.46	.865
10%	P4, D3	\$1.44	\$1.42	.072	\$2.29	\$2.26	.154
20%	P3, D4	\$1.34	\$1.30	< .001	\$2.13	\$2.07	< .001
30%	P3, D4	\$1.24	\$1.18	< .001	\$1.97	\$1.87	< .001
40%	P3, D4	\$1.14	\$1.06	< .001	\$1.81	\$1.68	< .001
50%	P2, D4	\$1.05	\$0.94	< .001	\$1.65	\$1.48	< .001

While this assumption is unlikely to hold in practice (agents would presumably adjust their behavior in response to different destruction rates), it allows us to isolate the mechanical effect of litigation costs on both the equilibrium and on the consequences of participant selection. The Nash Equilibrium shifts across destruction rates favoring more conservative choices as impasse gets costlier, and outcomes are increasingly driven by reaching any agreement.

Principal's Choice of Agent

We now look at the choice the Principals made. As predicted, Plaintiffs generally preferred an Agent who provided them with a high forecast: 48% of Principals selected an Agent who was among the two most favorable Agents (quintile 4 or 5). Indeed, the average Agent chosen comes from a quintile that is significantly higher than the ex-post optimal choice of 3 ($t(238) = 6.56, p < .001$). Defendants' decision mirrored that of the Plaintiffs: 71% of clients chose a lawyer who provided an estimate that was more favorable than the 4th quintile ex-post optimal choice – that is from either quintiles 1, 2, or 3 ($t(234) = -15.66, p < .001$). As predicted, we see that Principals on both sides on average choose an Agent who is more favorable to them than the ex-post optimal strategy suggests.⁹

Returning to Figure 2, we can now look at the subset of Agents who are in the ex-post

⁹ We show the distribution of choices for Plaintiffs and Defendants in the Appendix, Figure A2.

optimal quintiles (Optimal Agents, shown in green). We see that their beliefs about the judge's ruling are less polarized than any of the other comparisons, and in fact do not differ significantly from one another (\$31,969 for Plaintiffs and \$29,373 for Defendants, $t(79) = 1.00$, $p = .318$). Thus, one of the benefits seems to be that these less aggressive agents are better calibrated about the value of each party's walkaway option and, therefore, have a clearer sense of what a fair outcome would be.

The Agent-Selection Dilemma

So far, we have shown that Principals could increase their earnings if they *both* refrained from engaging in agent selection and that, conditional on selecting Agents, they tend to choose an Agent who makes (and believes) grander claims of the outcome they can achieve than the ex-post optimal Agent choice. In this section, we show one potential reason for why people may nonetheless choose to engage in agent selection in the real world. We propose that agent selection may be the result of a prisoner's dilemma, where random selection (i.e., hiring the first Agent one meets) represents a choice of cooperation and agent selection represents a choice of defection. We construct a 3×3 payoff table for a game in which each party decides whether to select an Agent at random from their set of five, whether to engage in the (biased) agent-selection process, or whether to choose the ex post optimal Agent (shown in Figure 4).

Looking first at a subset of actions, choosing randomly or engaging in agent selection, we observe a prisoner's dilemma-like pressure away from the former towards the latter. Specifically, the random-random pairing (top left) maximizes joint earnings, but both parties have an incentive to engage in agent selection on the margin. Therefore, choosing an Agent is advantageous when the other party selects at random (or, equivalently interpreted, hires the first Agent they encounter). This incentive to defect yields an equilibrium outcome of both parties engaging in agent selection (purple). It is conceivable that the choice of an aggressive Agent is optimal when the other party does not engage in selection. That is, participants may act as-if the other party were selecting randomly and fail to account for their (equally biased) selection. However, we find that selecting the ex-post optimal agent *conditional on both parties engaging in selection* would

also lead to better outcomes against a counterpart who selected at random (the 3x3 extension of Figure 4). Indeed, this ex post optimal agent choice outperforms random selection, suggesting that the problem is not selection itself but whom Principals choose as their representatives.

Figure 4

In a 2 x 2 game of choosing at random or selecting an Agent, the pure strategy Nash equilibrium is to engage in agent selection (Study 1). The Optimal Agent rows and columns represent an extension of the game and show that choosing an aggressive Agent also leads to worse outcomes when playing an opponent who does not engage in selection.

Defendant's Strategy	Random Agent	(\$1.97, \$1.24) Impasse: 65%	(\$1.83, \$1.27) Impasse: 75%	(\$1.86, \$1.32) Impasse: 68%
	Selected Agent	(\$2.02, \$1.15) Impasse: 69%	(\$1.87, \$1.18) Impasse: 79%	(\$1.89, \$1.22) Impasse: 74%
	Optimal Agent	(\$2.14, \$1.24) Impasse: 52%	(\$1.96, \$1.25) Impasse: 66%	(\$2.03, \$1.34) Impasse: 53%
		Random Agent	Selected Agent	Optimal Agent
		Plaintiff's Strategy		

Discussion

In Study 1, we present evidence that the agent-selection process, which precedes instances of principal-agent relationships, can contribute to impasse. We showed that Principals select

Agents who present overly optimistic forecasts about the outcome of the case, are more biased than Agents overall in their assessment of a neutral judge's ruling, and subsequently make offers in the ultimatum settlement task that are more aggressive than optimal. Notably, this is not the result of misaligned incentives or Agents deceiving their potential Principals. Rather, Principals select aggressive Agents who genuinely hold more optimistic beliefs and act upon them. As a result, impasse rates increase and both Principals are worse off than if they unilaterally deviated and selected a less aggressive Agent instead, or if they selected at random (i.e., the first Agent they encountered). Such random selection, however, is off the equilibrium path and hence would not be sustainable without a policy intervention.

Study 2: Agent-Selection vs. Self-Representation

One possible explanation for why Principals in Study 1 chose overly aggressive Agents is that the Principals themselves may already be severely polarized—more so than the average potential Agent—and they simply select Agents who share their extreme views. In this account, selection merely “passes on” the Principal's polarization to the Agents that represent them. Alternatively, it is possible that the agent-selection process causes Principals to observe and select Agents who hold viewpoints that are even more extreme than their own initially-held beliefs. They may then engage in a form of wishful thinking, selecting this more extreme Agent, who is giving them “good news” about the outcome that can be expected (Eil & Rao, 2011). Principals may then even internalize this favorable forecast and become more extreme in their own beliefs about what a fair outcome would be.

In line with this second account, we hypothesize that agent selection increases polarization relative to a case of Principals representing themselves. We test this prediction by experimentally manipulating *representation*: participants either negotiate the settlement on their own behalf or select an Agent and then that Agent negotiates on their behalf.

Experimental Design

The design follows closely that of Study 1 and was also preregistered on AsPredicted.¹⁰ We recruited participants via a different online labor market, Prolific Academic. Participants were presented with the identical personal injury lawsuit and assigned to either the side of the Plaintiff or the Defendant. We extend the first stage by including an additional Representation manipulation: participants either compete to be lawyers for the Plaintiff or the Defendant (Agent-Representation condition), or they engage in the ultimatum settlement negotiation on their own behalf (Self-Representation condition). While Agents who compete to be hired again report an amount to the Principals in an effort to be hired, those acting on their own behalf report what they believe the outcome of the case will be.

We make three new predictions: first, we predict that Selected Agents will have more polarized beliefs about what a fair outcome would be than Principals representing themselves. As a direct result, we expect that individuals negotiating on their own behalf will achieve lower impasse rates and better outcomes than individuals who are represented by their Selected Agents. Moreover, we expect no difference in beliefs about a fair outcome between Principals representing themselves and the average Agent in the pool of potential agents for their respective side. Second, we predict that Agents anticipate this demand for favorable information from the Principals. As a result, they will communicate an outcome to Principals that is more favorable than what the Principals representing themselves report that they can achieve. Third, we predict that observing these inflated forecasts as part of the agent-selection process will cause the Principal to become more polarized in what they think is fair outcome of the case, relative to those representing themselves in the negotiation. To test the third new hypothesis, we made a change to the order in which information is presented to participants in Stage 2: we now ask Principals to estimate the judge's ruling *after* observing the signals from the prospective Agents, rather than before.¹¹

¹⁰ https://aspredicted.org/blind.php?x=1L7_6PX

¹¹ In addition, we made a few minor changes from Study 1: participants now read the detailed instructions about the ultimatum settlement game prior to forecasting the outcome of the case and report outcome measures in increments of \$500. For prospective agents, we removed the fixed \$1 bonus for getting hired and instead highlighted more that they

Finally, this experiment enables us to attempt to replicate our findings from Study 1 in a different sample of participants.

Results

We recruited 1034 participants for Stage 1 of the experiment. Of those, 91 did not pass the comprehension check and 111 provided an inconsistent offer in the ultimatum negotiation game. As preregistered, these participants were excluded, leaving us with 832 participants. The following day, we recruited 509 new participants for the role of Principals. Of those, 57 failed the comprehension check, leaving us with a final sample size of 452 Principals.

Negotiation Outcomes

To compute negotiation outcomes, we follow the same procedure as in the previous study. New in this experiment is the case of self-representation, which follows the same approach with respect to computing expected earnings: each self-representing participant on one side is pitted against each of the self-representing participants on the other side.

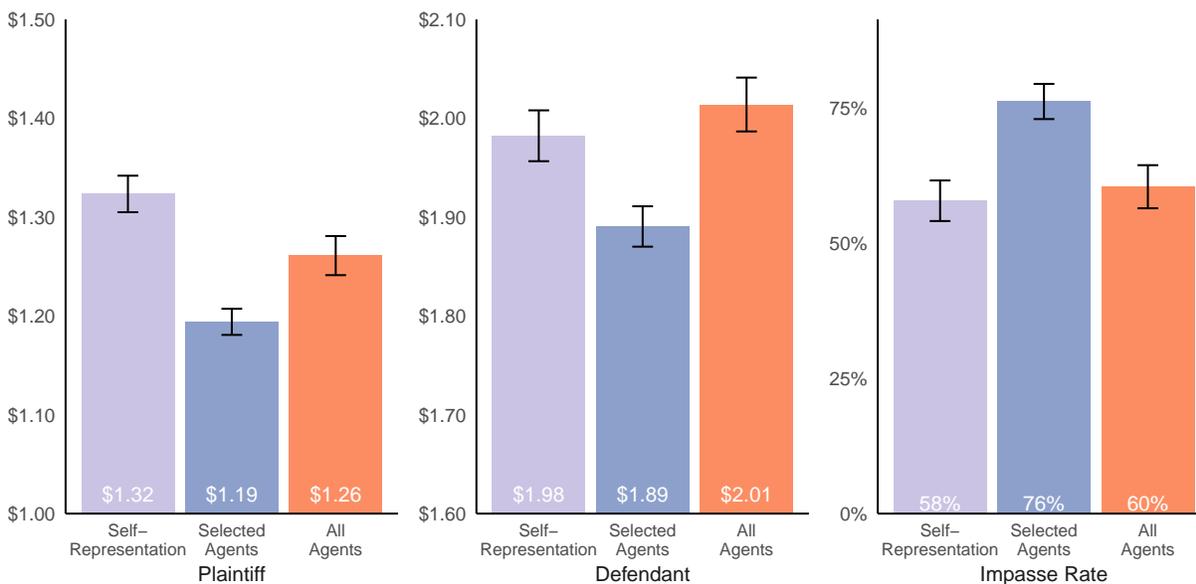
Figure 5 shows the earnings across the different groups of negotiators. Our main comparison is between participants representing themselves and the Selected Agents. We find that participants selected Agents who achieved worse outcomes than they themselves would have achieved when negotiating (both for Plaintiffs, $t(395) = 11.14$, $p < .001$, as well as for Defendants, $t(429) = 5.44$, $p < .001$). Also, replicating our results from Study 1, the agent-selection process again makes both parties worse off than if they selected an Agent at random ($t(399) = 5.49$, $p < .001$ for Plaintiffs and $t(415) = 7.11$, $p < .001$ for Defendants).¹² The right panel of Figure 5 shows the impasse rates across the three conditions. The impasse rate is lowest with Self-Representation at 58% and not significantly higher with All Agents at 60% ($t(830) = 1.07$, $p = .286$). Replicating the finding from Study 1, it is highest with Selected

would receive the (larger) bonus from the outcome of the negotiation only in the event that they are hired by someone. We also further emphasized that clients would make their hiring decision exclusively based on the forecast they submit. The exact materials are shown in the Supplementary Materials.

¹² We show the range covered by each quintile of Agents in the Appendix (Figure B1), along with the payoff matrix for each pairing of quintiles (Figure B3).

Figure 5

Negotiation outcomes in Study 2. The first two panels show average earnings for Plaintiffs and Defendants, respectively. The third panel shows the impasse rate. Self-representation outperforms agent selection, and randomly selecting from all Agents again outperforms Selected Agents, replicating Study 1. Agent selection increases impasse rates. Error bars show 95% confidence intervals.



Agents at 76% ($t(826) = 4.50, p < .001$ and $t(816) = 3.15, p = .002$, relative to Self-Representation and All Agents, respectively).

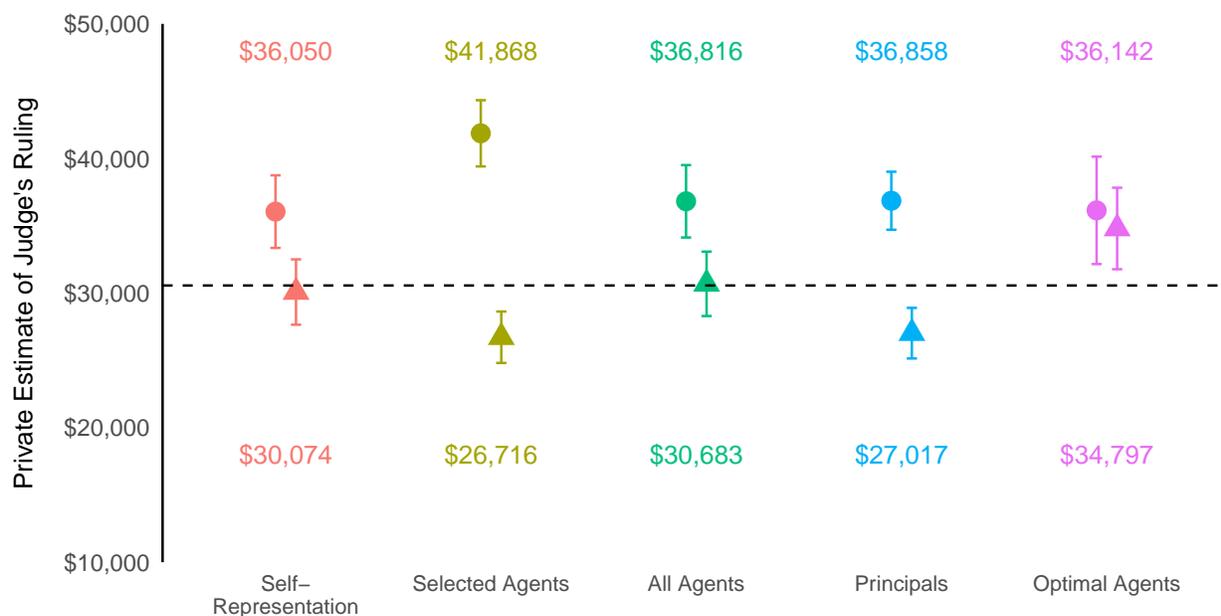
Beliefs About Fair Outcome

We next compare participants' beliefs about a fair outcome—their incentivized private estimates of what the neutral judge would decide (Figure 6). Our corresponding preregistered OLS regression specification includes a main effect for the role (Plaintiff or Defendant), whether they acted as an agent, and the interaction (Table 2). As predicted, we observe that Selected Agents hold more extreme beliefs than those negotiating on their own behalf (see the interaction term in Column 1). That is, the Principals are not merely passing on their polarized beliefs. Agent selection is leading to more polarized parties being at the bargaining table than if individuals negotiated on their own behalf.

In Column 2 of Table 2, we compare the beliefs of those representing themselves to those

Figure 6

Selected Agents in Study 2 are more polarized in their beliefs than participants negotiating on their own behalf. Moreover, Principals become more polarized after observing the proposed outcomes of the available Agents than either the Agents making those proposals or those negotiating on their own behalf. Circles show beliefs of those on the side of the Plaintiff, triangles show the side of the Defendant. The dashed line shows the judge's decision and error bars show 95% confidence intervals.



of Agents overall. Here, and as expected, the interaction effect is not significant. That is, beliefs do not differ merely as a result of acting as an Agent instead of on one's own behalf. Column 3 replicates the result from Study 1 that the beliefs of Selected Agents are more polarized than those of Agents overall. Finally, Column 4 compares the beliefs of those representing themselves and those of the Principals. These two groups differ only in that the Principals observed the forecasted outcomes of all available Agents prior to making their estimate. We find a non-significant increase in polarization after observing the Agents' signals ($p = 0.099$).

Agents' Signals

When Agents communicate the outcomes they think they can achieve to their prospective clients, they are incentivized to get hired rather than to be accurate. We predicted that Agents would anticipate the Principals' desire for good news and therefore "shade" the outcome they claim they can achieve, supplying favorable information relative to the comparable estimates of

Table 2

Beliefs about the fair outcome (Judge's Ruling) of different groups in Study 2. Participants who act on their own behalf are less polarized in their beliefs than Selected Agents (Column 1) and no different from Agents overall (Column 2). Selected Agents are again more polarized as Agents overall (Column 3). Contrary to our expectations, those observing proposed outcomes by potential Agents are not more polarized than those who did not do so because they represented themselves (Column 4).

	Self vs. Selected	Self vs. All	All vs. Selected	Self vs. Principals
Defendant	-5975.339*** (1708.384)	-5975.339** (1831.579)	-6132.602*** (1708.755)	-9840.141*** (1623.124)
Agent	5818.020*** (1757.125)	766.176 (1867.481)	5051.844** (1730.148)	
Defendant x Agent	-9176.014*** (2435.797)	-157.262 (2604.345)	-9018.752*** (2423.064)	
Self-Representation				-808.079 (1669.739)
Defendant x Self-Representation				3864.802+ (2338.720)
Constant	36 050.000*** (1237.771)	36 050.000*** (1327.030)	36 816.176*** (1212.674)	36 858.079*** (1140.079)
N	828	832	818	873

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

achievable outcomes made in the self-representation condition. We conduct an OLS regression with main effects for the role of Defendant and the Self-representation condition, as well as their interaction. Contrary to our expectations, we find no evidence of such shading by potential agents ($b = -596.80$, 95% CI [-6, 102.13, 4, 908.52], $t(828) = -0.21$, $p = .832$). This may explain why we do not observe an increase in belief-polarization among the Principals who have observed these projected outcomes. Notably, however, this suggests that a strategic motivation to present favorable information is not a necessary component of polarization resulting from agent selection—natural variation in beliefs among potential agents is sufficient to generate the predicted negotiation dynamics even in the absence of strategic shading.

Discussion

We found that selected agents were more polarized in their beliefs about a fair outcome and more aggressive in their bargaining than people negotiating on their own behalf. Further, self-representation did not yield greater belief polarization than bargaining between the average potential agent for each side. Therefore, in the process of agent selection, principals pick agents with more extreme beliefs in their favor, as opposed to beliefs that most closely resemble their own. Why do they do this? We find directional (but not statistically significant) evidence that Principals update their beliefs about the fair outcome after observing what Agents report they can achieve when negotiating on their behalf (a pattern we will return to in the next study). This holds even as these reported projections do not differ from what those reporting on their own behalf believe they can achieve, on average. This suggests that they put more weight on favorable projections and thus may persuade themselves that a better outcome is plausible, likely, and potentially even fair. As a consequence, the increased polarization of selected agents led them to achieve worse bargaining outcomes than the case of random agent selection or self-representation.

Study 3: Evolution of the Agent Pool

Thus far, the pool of available agents in our studies has been static. Principals received a representative sample of agents to choose from and selected agents with overly favorable views. In the real world, however, agents are subject to competitive pressures: if they fail to be hired by principals, they may be driven out of the market. In this experiment, we extend our findings by examining the consequences of how the pool of potential agents changes over time as a result of selection. We use an ecological lens ([Axelrod & Hamilton, 1981](#)) to examine a dynamic setting in which agents who fail to attract principals disappear from the market, whereas those who have been selected frequently will be more likely to be presented to principals. That is, agents not selected do not “survive,” and agents selected often “proliferate” in future rounds.

In this context, we examine whether principals, despite being fully informed about the selection process, continue to choose relatively optimistic agents or whether they moderate their choice over time. We predict that they will naively continue to prefer favorable agents from their

consideration set. Further, since that consideration set is becoming more extreme over time, we expect that polarization of selected agents increases for principals in later rounds. Notably, this might lead to such a skewed pool of surviving agents that even a principal trying to account for selection may not encounter an agent whose beliefs about the outside option correspond with the truth.

Finally, we increase the opportunity for motivated reasoning by introducing information asymmetry between the principals and the agents. Specifically, we provide principals with only the depositions of the plaintiff and the defendant. These documents provide the facts of the case from principals' perspectives. However, the potential agents also observe the expert depositions, witness statements, and driving records, as in the previous studies. In many real-world situations, there is a gap between principals and potential agents in how informed they are. This study introduces such information asymmetry to test whether our effects are replicable under these conditions.

Experimental Design

Our study took place across five rounds. In each round, we randomly assigned participants to the role of Plaintiff or Defendant. We introduced them to the personal injury lawsuit from the previous studies and presented them with the depositions of the Plaintiff and Defendant in the case. Participants then completed a four-item comprehension check (we dropped a question referring to information presented in one of the left-out documents) and excluded all participants who failed to answer the questions correctly on either the first or the second try. We targeted a sample of 100 participants per role per round, limited to those who passed the comprehension check. We used Quotas in the Qualtrics survey platform to target this number and ultimately ended up with a total of 1010 participants, including at least 100 for each role in each round.

We then presented participants with five Agents drawn from Study 2 and asked them to select one as their representative. In all rounds, we informed them that their Agents had had access to eight additional documents and based on this information engaged in the ultimatum settlement game (see Studies 1 and 2). Starting with Round 2, moreover, we informed them about

how the pool of Agents they are seeing was shaped by previous Principals' choices. Specifically, at the end of each round, we updated the pool of Agents by weighing them according to how often they were selected that round. For example, an Agent who was chosen twice in one round would be twice as likely to be drawn than an Agent who was chosen only once. Moreover, an Agent who was never selected would disappear from the pool. As a result, participants starting with Round 2 knew that their potential Agents had been selected at least once before. As in our previous studies, we presented Principals with one Agent from each quintile, based on the expected outcome communicated by the Agent. Therefore, the range of each quintile changed over time based on Principals' choices. We informed participants that this was based on 100 participants in each preceding round. In rounds when our actual sample exceeded 100 participants, we used the decisions of the first 100 participants in that role to determine the updated pool of Agents.

After participants observed the projected outcomes from five Agents and made their selection, we asked them to estimate the Judge's decision. As in previous studies, they could earn 50 cents if their estimate was within \$5,000 of the correct answer. The experiment then concluded with basic demographic questions. We preregistered our hypotheses, sample size, exclusion criteria, and analyses on AsPredicted.¹³

Results

We begin by replicating our primary result, comparing the outcomes of Selected Agents with those of All Agents. In Round 1 of this study, the pool of Agents is identical to that in Study 2. This test enables us to see if our primary result still holds now that we have introduced information asymmetry, albeit with a smaller sample. We find that the expected earnings in the first round do not differ between Selected and All Agents for Plaintiffs ($\Delta M = 0.00$, 95% CI $[-0.03, 0.04]$, $t(341) = 0.29$, $p = .776$) but Selected Agents lead to significantly lower earnings for Defendants ($\Delta M = 0.17$, 95% CI $[0.11, 0.22]$, $t(327) = 6.06$, $p < .001$). We replicate the polarization in the Principals' beliefs about a fair outcome among the Selected Agents (\$44,319 and \$32,168 for Plaintiffs and Defendants, respectively, $\Delta M = 12,151.13$, 95% CI

¹³ https://aspredicted.org/LJ9_J4Y

[7, 101.71, 17, 200.54], $t(207) = 4.74$, $p < .001$).¹⁴ As in previous studies, Selected Agents were more polarized in their beliefs about a fair outcome than the pool of Agents in general ($b = 7, 270.96$, 95% CI [1, 053.81, 13, 488.11], $t(668) = 2.30$, $p = .022$), with costly consequences for Defendants. Next, we turn our attention to the main focus of this study: the dynamics unfolding over time.

Table 3

The outcome dynamics in Study 3. Over time, the projected outcomes observed by the Principals are increasingly polarized (Column 1), and so are their beliefs about a fair outcome of the case (Column 2).

	Observed Proposed Outcome	Beliefs About Fair Outcome
Plaintiff	17 194.022*** (413.718)	15 048.313*** (1759.752)
Round	-330.158** (124.246)	321.159 (513.997)
Plaintiff x Round	6183.920*** (170.938)	2557.893*** (721.878)
Constant	28 971.173*** (266.892)	32 079.523*** (1257.772)
N	5050	1010
Clustered SE	Participant	No

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

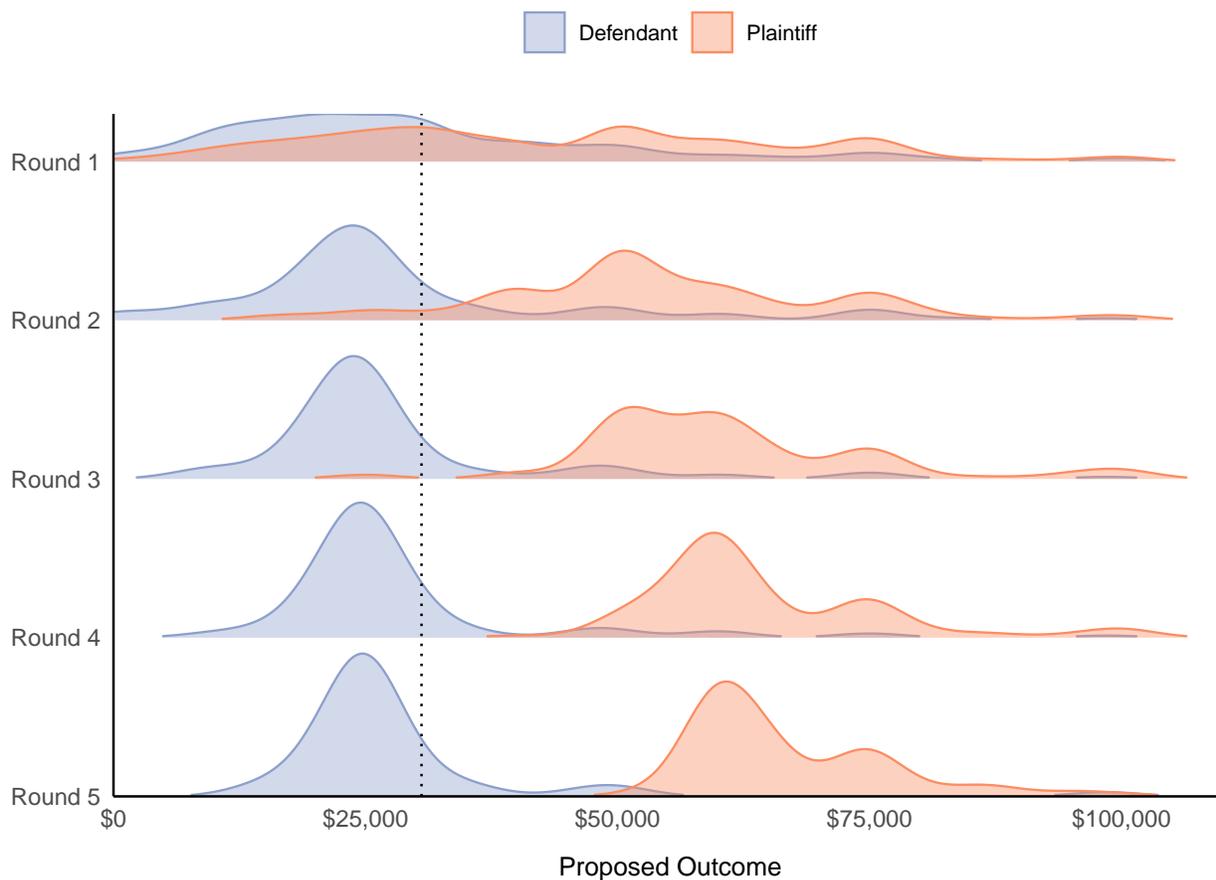
Recall that Principals' choices in each round determine the composition of Agents in the following round. We predicted that Principals choosing overly favorable Agents would lead to the disappearance of "moderate" Agents, thus increasing the gap between the projected outcomes observed by the Plaintiffs and the Defendants. We visualize this increasing polarization in Figure 7, showing the distributions of what the available Agents claimed they could achieve if selected. In Round 1, there is considerable overlap in the projected outcomes that the Plaintiffs and Defendants observe from their potential Agents and the range covers the judge's ruling. By Round 2, however, Plaintiffs are already unlikely to see an Agent proposing an outcome that is not

¹⁴ Since we use the Agents from Study 2, the analysis for All Agents is equivalent to that reported earlier. Recall that we also observed polarization there.

above the judge’s “fair” decision. By Round 3, the most pessimistic Agent of one party is very unlikely to overlap with the most optimistic Agent available to the other party. By round five, even the most pessimistic Agents for each side no longer overlap meaningfully. We show the corresponding statistical analyses in Column 1 of Table 3. Specifically, we conduct a linear regression where the outcome measure is the outcomes proposed by the Agents seen by the Principal, and predictors are the side of the dispute, the round, and their interaction. We predicted, and find, a significant interaction between role and round, which captures the increasing polarization between what the parties observe ($p < 0.001$).

Figure 7

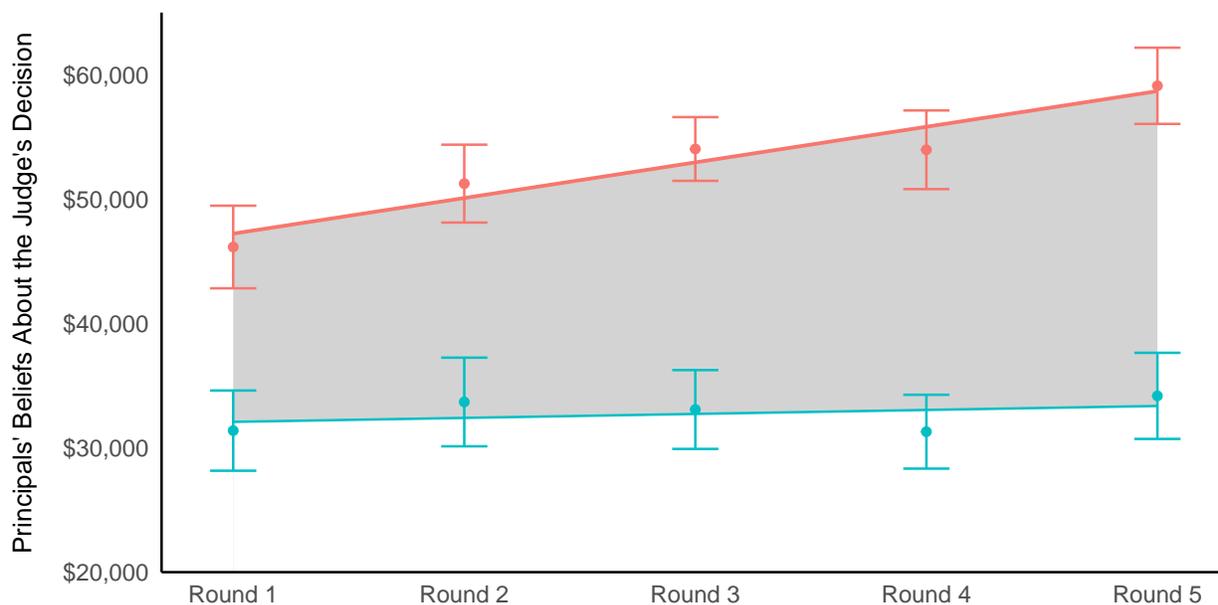
Distribution of proposed outcomes by round in Study 3. In Round 1, Defendant (blue) and Plaintiff (orange) proposals overlap substantially near the judge’s ruling (dashed line). As moderate agents are eliminated through selection, the distributions shift apart: by Round 3 there is almost no overlap, and by Round 5 the two sides propose starkly different outcomes.



Next, we examine how Principals' beliefs about a fair outcome (the judge's decision) differed by round as a consequence of the pool of potential Agents becoming more polarized. We expected that Principals in later rounds (relative to those in earlier rounds) would end up more polarized in their private beliefs about what a fair outcome would be, as a consequence of having "interacted with" a pool of potential Agents who are more polarized. This would occur because the Principal is receiving communication from five potential Agents regarding what they believe they can achieve on the Principal's behalf, and this set becomes more extreme in later rounds. If such influence is occurring, then, as the Agents get more polarized, so should the beliefs of the Principals who select among them. Figure 8 shows that this is indeed the case. In our data, this polarization appears to be primarily driven by the Plaintiffs (the corresponding regression is shown in Column 2 of Table 3). As we noted earlier, Principals are already polarized in the first round, with Plaintiffs expecting a ruling that is almost 50% higher (and thus more favorable to them) than the Defendants. By Round 5, the gap between the two Principals nearly doubles.

Figure 8

Evolution of Principals' beliefs about the fair outcome in Study 3. We again find that Principals are polarized in their beliefs about the neutral judge's ruling in Round 1. Moreover, this difference increases over time as Plaintiffs become increasingly optimistic about how favorable an outcome they can receive if they do not settle the case. Error bars show 95% confidence intervals.



Finally, we look at how the increasingly polarized Agent pool and the increasingly polarized beliefs of the Principals affect the outcomes for the two parties. Figure 9 shows the impasse rates over the five rounds, with bootstrap confidence intervals computed as in Studies 1 and 2. Not surprisingly, we find that the Selected Agents are less likely to reach a settlement agreement with each round ($b = -0.03$, 95% CI $[-0.04, -0.01]$, $t(499) = -4.27$, $p < .001$). The impasse rate among Selected Agents rises from 72% in Round 1 to 82% in Round 5. When we look separately at the earnings of Plaintiffs and Defendants, we see that the cost of this increasing impasse rate in our experiment comes entirely at the expense of the Plaintiffs, who lose about 10% of their potential earnings by the fifth round ($b = -0.03$, 95% CI $[-0.04, -0.02]$, $t(507) = -10.06$, $p < .001$). Defendants see no decrease in earnings over time ($b = 0.00$, 95% CI $[-0.01, 0.01]$, $t(499) = -0.11$, $p = .915$). This suggests that Plaintiffs' increasingly "optimistic" Agents are more likely to turn down offers that are better for their clients than the judge's ruling. Whereas Defendants select Agents who reject unfavorable settlement offers—offers that they now should reject.

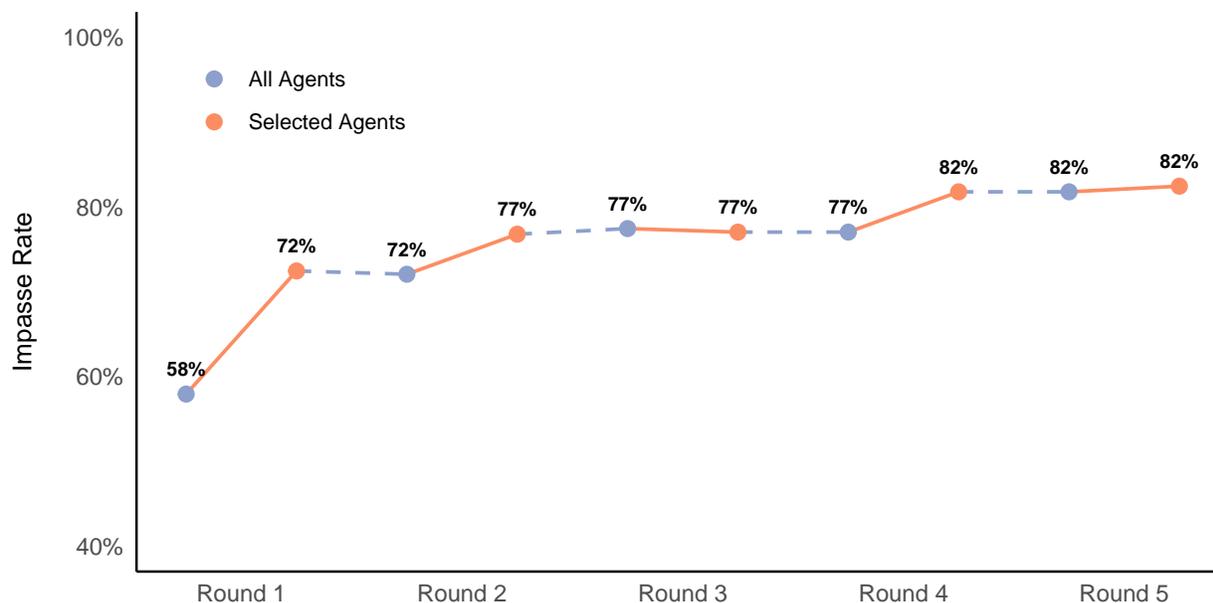
Note that the sample of Selected Agents changes in two ways over time: the mean in expected outcomes becomes more polarized, because Principals choose favorable agents, and the variance in beliefs becomes smaller because the least favorable available Agents disappear. To see which of these components is driving the increase in impasse rates, we simulate a counterfactual in which we keep the mean (variance) in Round 1 fixed across rounds and simulate data using the real variance (mean) in subsequent rounds. This decomposition, visualized in Figure C1 in the Appendix, shows that the result is driven by the mean shift. Participants persist in selecting optimistic Agents and fail to account for the selection that had already occurred.

Discussion

We show that when selection choices affect the survival and proliferation of agents across rounds, the pool of potential agents becomes increasingly extreme over time. Principals in each period select agents who are more favorable to their side than the median agent. Subsequent principals then fail to account for the dynamics of others having selected favorable agents in

Figure 9

Impasse rates in Study 3 over time. As the agent pool becomes more polarized through selection, Selected Agents (orange) become increasingly likely to reach impasse compared to All Agents (blue).



previous rounds. As a result, the pool of available agents continues to diverge between the two parties. By the third period, there is little overlap between the projected outcomes of even the most pessimistic agents observed by the plaintiff and the defendant and settlement negotiations by selected agents increasingly end in impasse. These dynamics resemble the survival of overconfident traders in markets (Benos, 1998). This divergence ultimately also influences the beliefs of principals themselves. Principals in later rounds—who observe the proposed outcomes from their consideration set of potential agents—become increasingly polarized about what a “fair” outcome would be, as the pool of agents they observe becomes more extreme.

Our effect is primarily driven by the beliefs and actions of the plaintiffs. This may be because of the asymmetric nature of the case: the judge’s decision of \$30,560 is below the mid-point of \$50,000—where settlements could range from \$0 to \$100,000—and the case materials include the plaintiff’s annual salary (\$20,000) as well as the medical costs (\$10,000). As a result, proposed outcomes that suggest the plaintiff should not offer any payment may be unrealistic and hence explain why the most favorable agents for the defendant vanish from the

pool. By contrast, the most optimistic agents for the plaintiff survive through the end. Notably, it is also only the plaintiffs whose earnings suffer over time: defendants remain better off by accepting an impasse and taking the case to trial than accepting increasingly excessive settlement offers from the plaintiffs' agents.

Study 4: Interactive Negotiation

Our previous studies used a stylized setting in which bargaining consisted of making ultimatum offers and setting rejection thresholds. This enabled us to simulate counterfactuals and identify optimal agent quintiles. However, the opportunity for counterfactual simulation has come at the expense of realism in the nature of the negotiation itself, which could limit the generalizability of the results (Alekseev et al., 2017; Galizzi & Navarro-Martinez, 2019; Levitt & List, 2007; Wang & Navarro-Martinez, 2023). Therefore, in this study we introduce a new design that resembles a more naturalistic negotiation. We preregistered our hypotheses, sample size, exclusion criteria, and analyses on AsPredicted.¹⁵

Experimental Design

We developed a negotiation platform that allows participants to engage in a live, text-based, open-ended bargaining interaction after being matched with another participant over the internet. Both parties could send justifications for their positions and make, counter, accept, and reject offers. The negotiation task involved a business setting centered on a commercial sale, with participants assigned to the roles of Buyer or Seller. Participants received a rich set of stimuli providing them with information they could use to bargain, as well as materially relevant but noisy information about the value of their outside option. Impasse in the negotiation occurred when the participants could not reach an agreement by the deadline and hurt total joint value for the two parties. But as with our previous studies, each side had an incentive to turn down offers they thought were worse than their outside option.

In line with our previous studies, participants reported the expected outcome prior to engaging in the negotiation. A new group of Principals then selected an Agent based on the

¹⁵ <https://aspredicted.org/9cg86a.pdf>

reported expected outcome. As in prior studies, we examine whether Principals chose someone whose expected outcome is favorable to their side. We then tested our central prediction: impasse rates increase (and hence joint earnings decrease) with two-sided agent selection relative to no agent selection (i.e., random selection). Using open, live negotiation precluded us from simulating counterfactuals; therefore, in this study we cannot compare agent selection to an ex post optimal benchmark, as we did in previous studies. Our primary outcome therefore is the impasse rate as a source of value destruction.

Agents

We recruited 761 participants via Prolific for the role of Agents. We randomly assigned them to the role of “Buyer” and “Seller” and presented them with materials for a distributive negotiation resembling those used for MBA courses. Specifically, participants learned some information common to both sides that provided background information (e.g., the object of negotiation), along with some information that was private to their role (e.g., information about their outside option). The context provided richness for discussion and could be used to justify offers. Notably, distributive negotiations resembled the incentives in our stylized ultimatum negotiation game, in that both sides would prefer a deal over their outside option, but the outside option is preferable to a bad deal. However, this negotiation setting required us to introduce private information, and therefore the two parties may differ in their assessments of an outcome as a result not only of biased interpretation of the information, but also because they did not receive exactly the same information. Within a given role, however, beliefs can still vary because people may differ in their (perceived) negotiation abilities and in their assessment of their side’s bargaining power. To anchor their beliefs and introduce further experimentally-induced variation, we provided each side with a consultant’s estimate of their outside option. Participants were randomly assigned to an “optimistic” estimate or a “pessimistic” estimate. The true value of the outside option was in the middle and the signal thus conveyed accurate information across participants.

Participants were further informed about the procedure of the negotiation. They would be

paired with another Prolific participant in real-time for a text-based negotiation lasting for up to seven minutes during which they could message each other and propose offers. If they reached an agreement, their bonus was determined as follows: Buyers earned 10 cents for every \$100,000 below a price of \$4.0m, and Sellers earned 10 cents for every \$100,000 above a price of \$1.5m. Thus, participants had an incentive to get a good deal. If they could not reach an agreement, the price was determined by an outside option that differed for each role: \$2.5m for Buyers and \$3.0m for Sellers, corresponding to a \$1 bonus for each side. As noted, both sides got a signal about the value in the form of a consultant's estimate that was right on average. We assigned each role further to an Optimistic and a Pessimistic condition, such that Buyers received a signal of either \$2.0m (Optimistic) or \$3.0m (Pessimistic), and Sellers received a signal of either \$3.5m (Optimistic) or \$2.5m (Pessimistic). Note that any agreement between \$1.5m and \$4.0m was possible because participants did not know for sure what the value of their outside option was, and any such outcome would lead to a non-negative bonus and joint earnings of \$2.50. In case of an impasse, joint earnings were reduced by 20% to \$2.00, leading to less value destruction than in our previous studies.

After participants read the general and private information, they answered a three-item comprehension check, with each option having four potential answers. Specifically, they were asked about the object of negotiation, what happened in the event they could not reach an agreement, and what the consultant's estimate of the outside option was. Participants had two chances to answer the comprehension check correctly and proceed to a "preparation screen" helping them plan their strategy for the conversation. On that screen, we showed them both the common and the role-specific information about the negotiation (left side) and asked them to respond to open-ended questions to prepare (right side): participants in both roles wrote about (i) their target price (described as an ambitious outcome they aim to achieve), (ii) how they might reasonably justify this price to the counterpart, and (iii) at what price they would walk away and take the outside offer. Finally, and importantly, they answered what they predicted the outcome of the negotiation to be, selecting one of six multiple choice options ranging from \$1.5m to \$4.0m in

increments of \$500,000.

After completing this preparation, participants entered a waiting room in which they were matched with another available participant in the opposing role who had reached the same stage of the experiment. Participants who could not be matched within ten minutes due to either technical issues or a lack of available negotiation partners advance to the demographic questions and receive payment equivalent to reaching impasse. Those who were matched then proceeded to the negotiation screen (see Figure 10, and see the Supplementary Materials for all screens shown to participants) where they interacted for up to seven minutes. After the negotiation concluded, either with an agreement or with impasse, participants completed standard demographic questions and reported the extent to which they had prior experience engaging in negotiations.

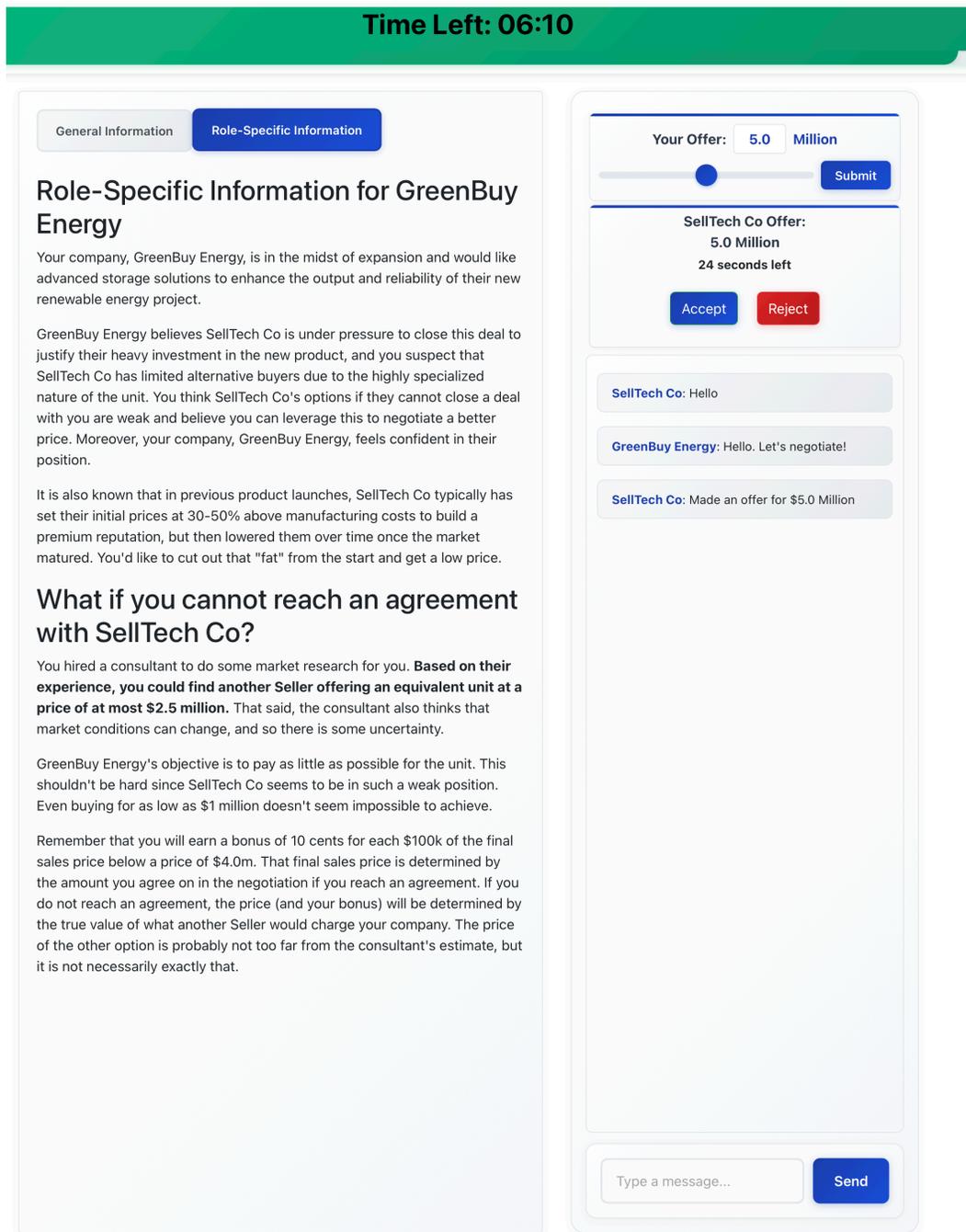
Principals

We then recruited a new group of 599 participants via Prolific for the role of Principal. We randomly assigned them to the side of Buyer and Seller and presented them with the identical information as the Agents had seen, except that we did not give them the consultant's estimate of the outside option. These participants also had to answer three items from a comprehension check correctly in order to proceed to the main task of the study. We then asked them to select an Agent based on one of the six possible outcomes they anticipated for the negotiation. Principals were incentivized with the same bonus system as the Agents and were informed that Agents had received the same outcome-based incentive. We further informed them that they would be matched with an Agent chosen by another participant in the opposing role, thus reflecting a two-sided selection task.

Results

For the first stage, we preregistered a series of exclusion criteria and targeted a sample size of 300 pairs after exclusions. Specifically, we excluded participants who failed the comprehension check after the second attempt (58), did not complete the preparation questions (50), were not successfully matched with a partner (15), or who were matched but at least one side did not take any action during the chat, such as sending a message, an offer, or making an accept or rejection

Figure 10
 Screenshot of the negotiation interface in Study 4. The left panel displays case information, including general and role-specific details. The right panel contains a real-time chat window where participants exchange messages and submit offers. A timer at the top counts down from seven minutes.

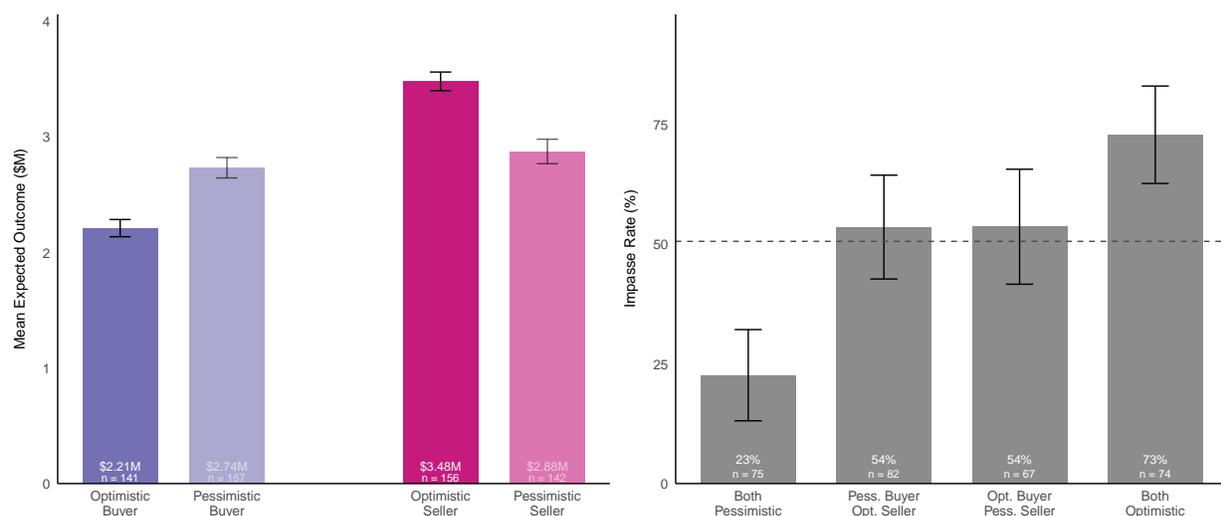


decision – all suggesting they were inattentive or absent during the negotiation (12), as well as pairs in which at least one party received a negative bonus, indicating an agreement price outside the plausible range (15). This leaves us with a total of 596 participants in 298 complete pairs.

Agents' beliefs about the expected outcome of the negotiation differed by role and by the signal they received about the value of the outside option (see left panel of Figure 11). Buyers who received a pessimistic estimate predicted that they would have to pay more than those who received an optimistic estimate (\$2.74M vs. \$2.21M; $t(296) = -8.76, p < .001$); Sellers who received a pessimistic estimate predicted that they would get less than those who received an optimistic estimate (\$2.88M vs. \$3.48M; $t(296) = 9.04, p < .001$). These differences in beliefs then translated into different impasse rates (right panel of Figure 11). While pairs overall reach a deal during the negotiation 49% of the time, this increases to 77% when both sides receive pessimistic forecasts for their respective position ($t(296) = 5.91, p < .001$), and decreases to 27% when their forecasts are optimistic for their respective positions ($t(296) = -4.56, p < .001$).

Figure 11

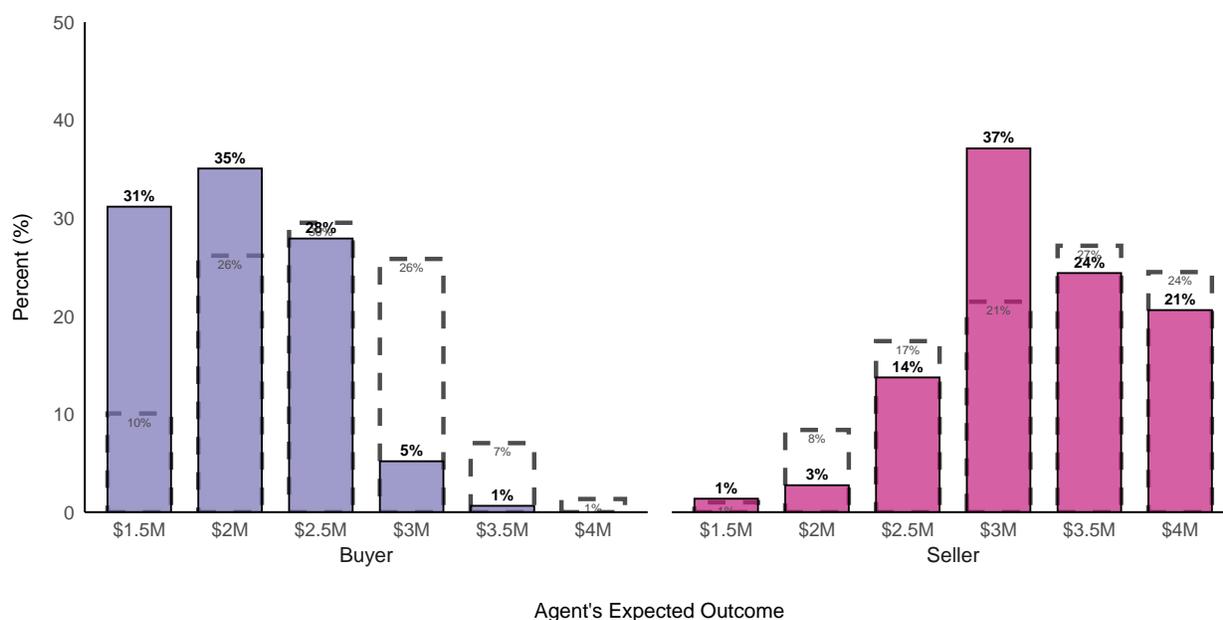
Left panel: Average expected negotiation outcomes by role and information valence in Study 4. Agents differed in their beliefs about what they could achieve as a result of getting an optimistic or pessimistic signal about the value of their outside option. Right panel: Impasse rates by valence combination. Impasse rates were lowest when both Agents received a pessimistic signal about their outside options, and highest when it was optimistic for both. Dashed line shows overall average impasse rate. Error bars show 95% confidence intervals.



We now turn to Stage 2 and look at whom the Principals chose as their representatives. Figure 12 compares the distribution of Agents selected by Principals to the distribution of available Agents from Stage 1. Buyer Principals selected Agents with a mean expected outcome of \$2.05M, compared to the available pool mean of \$2.49M ($\Delta M = -0.44$, 95% CI $[-0.53, -0.36]$, $t(604) = -10.46$, $p < .001$). Seller Principals, however, selected Agents with a mean expected outcome of \$3.21M, versus an available pool mean of \$3.19M ($\Delta M = 0.02$, 95% CI $[-0.08, 0.12]$, $t(587) = 0.33$, $p = .740$), which is not significantly higher. We report the payoff matrix for each pairing in the Appendix (Figure D1), although note that it is significantly less populated than previous our studies due to the inability to simulate counterfactual negotiations in this study.

Figure 12

Distribution of available Agents (dashed outline, from Stage 1) and Agents selected by Principals (solid bars, from Stage 2), by role in Study 4. Percentages are shown for selected (bold) and available (grey) distributions.



To assess the consequences of selection, we compare outcomes under two scenarios. The *Random Agent* baseline uses the actual Stage 1 outcomes, where Agents were randomly paired. The *Selected Agent* scenario simulates the expected outcomes when each Principal's chosen Agent is paired against an Agent drawn from the opposing side's selection distribution, weighted by the

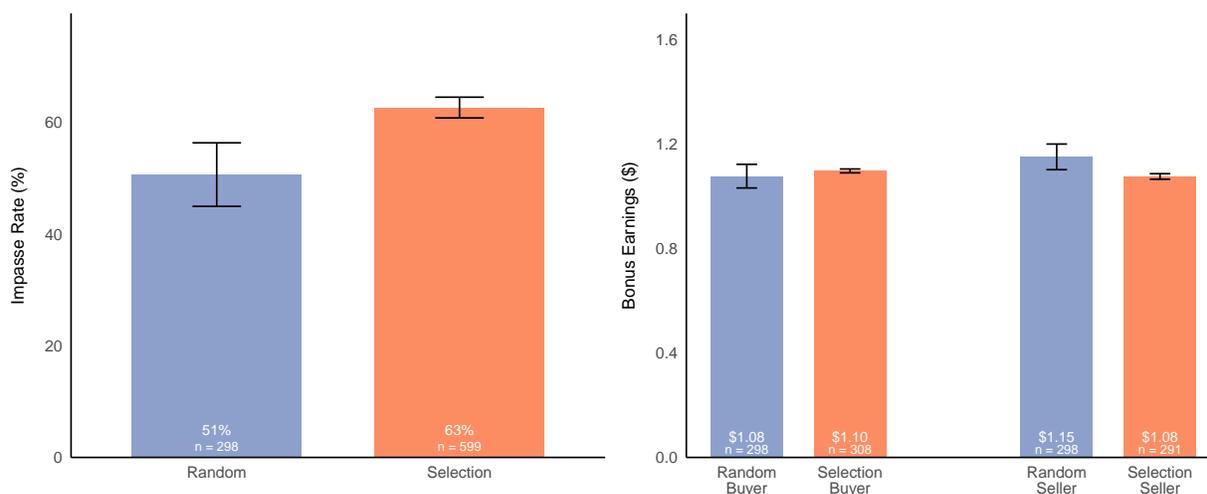
observed payoff matrix. Specifically, we hold fixed the choice of the focal Principal, then calculate the expected earnings for each of the six types of Agents they could have been matched with, then weigh those outcomes by the probability that such an Agent was selected by the Principals in the opposing role. We calculate the impasse rate similarly and present our main result in Figure 13.

As is evident in the left panel, impasse rates increased from 51% under Random matching to 63% with Selection ($\Delta M = -0.12$, 95% CI $[-0.18, -0.06]$, $t(604) = -4.07$, $p < .001$).

Looking at earnings (right panel), Buyers capture a smaller share of the bargaining surplus, and their earnings do not differ significantly with and without selection ($\Delta M = 0.02$, 95% CI $[-0.02, 0.06]$, $t(604) = 0.87$, $p = .384$). Sellers, on the other hand, see their earnings decline as a result of impasses (from \$1.15 with random pairings to \$1.08 with selection, $\Delta M = -0.07$, 95% CI $[-0.13, -0.02]$, $t(587) = -2.94$, $p = .003$). Thus, selection again led to worse joint outcomes, with the cost here falling entirely on the Sellers.

Figure 13

Negotiation outcomes in Study 4. The left panel shows impasse rates and the right panel shows mean bonus earnings under Random matching (blue; actual Stage 1 outcomes) and Selection (orange; simulated from Principals' choices). Error bars show 95% confidence intervals.



Discussion

Study 4 extends our investigation of agent selection to an interactive negotiation setting, departing from the ultimatum game and legal environment of the previous studies. In ultimatum

bargaining, selecting an agent who makes a more extreme offer weakly increases impasse rates, leading to a higher rate of rejection. In this richer negotiation setting, an aggressive opening position might serve as an anchor that steers the negotiation in a favorable direction through counteroffers and back-and-forth exchange (“anchor high and justify,” as is often advised to negotiation students). If anchoring effects were dominant, agent selection could benefit rather than harm principals. Our results indicate otherwise (consistent with work cleverly titled “Starting high and ending with nothing,” Schweinsberg et al. (2012)). Even in an environment where both parties can send messages, justify their positions, and make multiple offers over a seven-minute interaction, agent selection increases impasse rates. Principals on the Buyer’s side disproportionately choose agents with more optimistic projections about the outcome of the negotiation, and two-sided selection again increases impasse rates relative to random pairing. Notably, the cost of impasse falls asymmetrically in this study: Sellers, who achieve more favorable deal terms on average when agreements are reached, bear the brunt of lost earnings from increased impasse. Buyers’ earnings remain roughly constant because deals that do occur heavily favor the Seller, meaning impasse is comparatively less costly for the Buyer under the parameters of this negotiation. As with the differences in outcomes for Plaintiffs and Defendants in our previous studies, we do not propose that this is a general effect: changes in the negotiation materials could strengthen or weaken one side’s position, and thus change how the surplus from reaching agreement is shared. Similarly, a higher share of value destruction from impasse could make both sides strictly worse off.

These findings suggest that the agent-selection dynamic documented in the earlier studies is not an artifact of the ultimatum game’s binary accept-or-reject structure. Instead, the core mechanism—principals selecting agents whose beliefs are more favorable to their position, leading to more polarized pairings at the bargaining table—appears to operate across different negotiation structures, communication modes, and domains. While the stylized studies made possible identification of optimal strategies, Study 4 demonstrates that the fundamental pattern persists in a setting that more closely approximates real-world bargaining interactions.

General Discussion

Many negotiations involve agents bargaining on behalf of others. While considerable research has studied agency problems—whether the interests of the agent are aligned with the interests of the principal they represent—the role that agent selection plays in determining bargaining outcomes has received little scholarly attention. In this paper, we examined whom principals choose to represent them in a series of pre-registered distributive bargaining experiments and found that they systematically select overly aggressive agents.

Leveraging a novel design that allows us to calculate the counterfactual outcomes if different agents had been selected, we show that principals choose agents who provide them with a forecast of the outcome that is more favorable to their position than does the ex-post optimal agent. The agents who are selected hold more polarized beliefs about what a fair outcome would be than do these ex-post optimal agents and all agents overall. We also compare actual agent selection against a strategy in which the principal hires the first agent she arbitrarily meets (which we refer to as random agent selection). This comparison reveals a prisoner's dilemma. In our data both parties would benefit from both implementing random agent selection (“cooperate, cooperate”) relative to both engaging in deliberate agent selection (“defect, defect”). But, conditional on their opposition implementing random agent selection, both parties can increase their earnings by engaging in deliberate agent selection. This may explain why agent selection is common in real world settings, despite evidence that both parties are worse off when both retain bargaining agents ([Ashenfelter et al., 2013](#)).

This paper makes several contributions to the literature on bargaining by isolating a fundamental, but thus far unexplored, aspect of bargaining through agents: agent selection. In practice, agents are often selected from a pool of potential agents, and we show that this dynamic can have downstream effects on bargaining performance and outcomes. We further introduce an experimental methodology for examining counterfactual outcomes in bargaining, which is particularly valuable as such counterfactuals are unobservable in the real world. Finally, our results shed light on when and how principals' agent-selection decisions may backfire by

increasing the incidence of costly impasse. This work raises important questions of whether the high rates of impasse observed in the real-world may be at least partially explained by whom principals are choosing to represent them—namely, overly aggressive agents.

Our work also connects with and contributes to an ongoing stream of research on how adviser confidence affects advisee judgment and decision-making. Advisers who express more confidence are more likely to be listened to (Sniezek & Van Swol, 2001) and advice delivered with confidence is less likely to be checked for accuracy (Sah et al., 2013). Advisees tend to assume that a more confident adviser makes more categorically correct judgments and is more knowledgeable (Price & Stone, 2004). Advisers may, therefore, express excessive confidence in the interest of being listening to, chosen, or held in high esteem (Radzevick & Moore, 2011; Van Zant, 2022). This miscalibration can cause overconfident advisers to lose credibility (Tenney et al., 2008), but only if there is feedback and information such that this adviser-miscalibration is clearly discernible from the perspective of the advisee (Sah et al., 2013). In our paper, overconfidence emerges in the belief extremity of selected agents, akin to the type of overconfidence referred to as overestimation by Moore and Healy (2008). We build on this past work by considering advisors in the form of agents, showing how a two-sided selection process, with natural variation in agent beliefs, is sufficient to generate situations in which the agents at a bargaining table overestimate their respective positions, which may cause negotiations to be overly contentious. In other words, principals choosing agents for bargaining may be selecting on overconfidence.

In this paper, we have focused on settings of aligned incentives in bargaining—from stylized ultimatum games (Studies 1–3) to live, open-ended negotiation (Study 4)—to isolate and illustrate a particular dynamic. It is plausible that in many contexts, there may be forces that exacerbate polarization through agent selection. In richer settings, for example, principals might reason that agents with more optimistic beliefs have more accurate private information or are better experienced negotiators, even as this need not be true. Further, our setting allowed little in the way of salesmanship by potential agents. In the real world, potential agents may engage in a rich set of behaviors to convey their confidence and optimism regarding the outcomes that a

principal can expect. Past work has suggests that individuals who are able to engage in self-delusion may be more likely to convince others (Schwardmann & van der Weele, 2019), and competitive environments may further select for such self-deluding individuals (Schwardmann et al., 2022). On the other hand, for bargaining that plays out over longer time horizons and with richer avenues for communication, one might observe agents that use misleading initial expectations to get selected and then, post-selection, use subsequent communication to temper the expectations of the principal. However, in our view, these effects would exist in addition to the agent-selection dynamic that we documented here, as we found that the selected agents in our study genuinely held biased, polarized beliefs about what a fair outcome would be. Future work may examine whether and how agents strategically misrepresent their position in order to get selected, and whether this ties them to the expectations they create. Recent work by Sadka et al. (2024) found that providing predictions for outcomes of labor dispute cases increases settlement rates, but only when the plaintiff is present and not when the prediction is only provided to their lawyer. This example suggests many more interesting interpersonal dynamics, in addition to potential conflicts of interest, present in real principal-agent interactions. While this paper highlights several fundamental aspects of agent selection and its bargaining consequences, we also believe there are many interesting elements still to explore in this domain and believe it could be a rich vein for future research.

It is also worth noting that our experimental design may understate the magnitude of the agent-selection effect relative to real-world settings. Prior psychological research documents “extremeness aversion,” whereby individuals tend to avoid selecting the most extreme option from a clearly ordered set, instead gravitating toward central options that appear to represent reasonable compromises. In our studies, principals view agents’ projected outcomes simultaneously in a clear, ordered format, which may encourage choosing a middle option. In many real-world agent-selection settings, however, options are not presented simultaneously in a clean rank order. Instead, principals often encounter potential agents sequentially or through less structured comparison processes, in which extremeness aversion may be less operative. To the extent that

this is the case, our experimental design may provide a conservative estimate of the tendency to select favorably extreme representatives.

The extremeness of selected representatives as a source of conflict extends beyond the domain of bargaining. Organizational and group processes may involve the implicit or explicit selection of a leader or spokesperson who holds a more extreme view than the average member of the group, in the interest of seeking strong advocacy. This dynamic may lead to excessive conflict when needing to coordinate across groups (e.g., within or between firms). Similarly, in the political domain, voters elect representatives on the basis of their claims of what they can achieve in office. Much like how our selected agents hold more polarized beliefs about the judge's ruling in the legal setting, elected officials and other elites may be more polarized in their policy attitudes than the pool of potential candidates in general, and have been shown to be more polarized than the voters whom they represent (Bafumi & Herron, 2010). Such polarization may then drive legislative gridlock (Hetherington & Rudolph, 2015).

An important question is whether these dynamics would emerge in practice where principals might learn from experience to avoid extreme agents. In many real-world principal-agent bargaining environments, however, such learning may be limited for multiple reasons. Principals rarely observe the counterfactual outcome of having selected a different agent and ex post attribution of impasse is often complex and ambiguous. When negotiations end in impasse, the failure is often attributed to the opposing party or to contextual constraints rather than to the agent selection decision itself. For example, in union-management negotiations, a union leader may attribute a breakdown to management's unwillingness to compromise, while management may make the same attribution in reverse. Without clear feedback about how alternative representatives would have performed on their behalf, it is difficult for principals to infer whether their own agent selection decisions are contributing to costly impasse. Additionally, many principal-agent bargaining settings occur infrequently—for instance, collective bargaining agreements are often negotiated only every few years. As a result, principals accumulate only a small number of observations from which to infer the consequences of their agent-selection

decisions. Limited feedback combined with noisy attribution may therefore impede the natural learning processes that might otherwise correct these dynamics. That said, in contexts in which counterfactual agents might be observed via vicarious learning or agent selection occurs repeatedly with clear performance feedback, astute principals may be able to identify the problem and select less aggressive agents.

Selecting agents to act on our behalf is at the core of many high-stakes decisions across a variety of managerial and consumer settings. For good reason, then, the principal-agent incentive misalignment problem has been established as a fundamental topic across the fields of management and economics. Our aim here was to show that the process of agent *selection* can be similarly fundamental in shaping the outcomes that end up materializing on the principals' behalf. While aligning incentives between principals and agents is amenable to intervention by existing policy tools (e.g., requiring disclosure of conflicts of interest or outright prohibiting the activity that may result therein, see [Lo et al., 2009](#)), the agent-selection process may be ripe with motivated reasoning and other established behavioral phenomena that may require new interventions to address. Given the substantial real-world costs resulting from unnecessary impasse, mitigating the agent-selection dilemma could substantially reduce deadweight loss and improve outcomes for all parties.

References

- Alekseev, A., Charness, G., & Gneezy, U. (2017). Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior & Organization*, *134*, 48–59. <https://doi.org/10.1016/j.jebo.2016.12.005>
- Amanatullah, E. T., & Morris, M. W. (2010). Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology*, *98*(2), 256–267. <https://doi.org/10.1037/a0017094>
- Amanatullah, E. T., & Tinsley, C. H. (2013). Punishing female negotiators for asserting too much...or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes*, *120*(1), 110–122. <https://doi.org/10.1016/j.obhdp.2012.03.006>
- Ashenfelter, O., Bloom, D. E., & Dahl, G. B. (2013). Lawyers as Agents of the Devil in a Prisoner's Dilemma Game. *Journal of Empirical Legal Studies*, *10*(3), 399–423. <https://doi.org/10.1111/jels.12014>

- Ashenfelter, O., & Currie, J. (1990). Negotiator Behavior and the Occurrence of Dispute. *The American Economic Review*, 80(2), 416–420. <https://www.jstor.org/stable/2006611>
- Aumann, R. J. (1976). Agreeing to Disagree. *The Annals of Statistics*, 4(6), 1236–1239. <https://www.jstor.org/stable/2958591>
- Axelrod, R., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science*, 211(4489), 1390–1396. <https://doi.org/10.1126/science.7466396>
- Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. (1995). Biased Judgments of Fairness in Bargaining. *The American Economic Review*, 85(5), 1337–1343. <https://www.jstor.org/stable/2950993>
- Bafumi, J., & Herron, M. C. (2010). Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress. *American Political Science Review*, 104(3), 519–542. <https://doi.org/10.1017/S0003055410000316>
- Bartling, B., & Fischbacher, U. (2012). Shifting the Blame: On Delegation and Responsibility. *The Review of Economic Studies*, 79(1), 67–87. <https://www.jstor.org/stable/41407045>
- Benos, A. V. (1998). Aggressiveness and survival of overconfident traders. *Journal of Financial Markets*, 1(3), 353–383. [https://doi.org/10.1016/S1386-4181\(97\)00010-4](https://doi.org/10.1016/S1386-4181(97)00010-4)
- Burson, K. A., Faro, D., & Rottenstreich, Y. (2010). ABCs of principal–agent interactions: Accurate predictions, biased processes, and contrasts between working and delegating. *Organizational Behavior and Human Decision Processes*, 113(1), 1–12. <https://doi.org/10.1016/j.obhdp.2010.05.002>
- Card, D. (1990). Strikes and Bargaining: A Survey of the Recent Empirical Literature. *The American Economic Review*, 80(2), 410–415. <https://www.jstor.org/stable/2006610>
- Choy, A. K., Hamman, J. R., King, R. R., & Weber, R. A. (2016). Delegated bargaining in a competitive agent market: An experimental study. *Journal of the Economic Science Association*, 2(1), 22–35. <https://doi.org/10.1007/s40881-015-0015-1>
- Coffman, L. (2011). Intermediation Reduces Punishment (and Reward). *American Economic Journal: Microeconomics*, 3(4), 77–106. <https://doi.org/10.1257/mic.3.4.77>
- Coffman, L., & Gotthard-Real, A. (2019). Moral Perceptions of Advised Actions. *Management Science*, 65(8), 3904–3927. <https://doi.org/10.1287/mnsc.2018.3134>
- Cummins, J. G., & Nyman, I. (2013). Yes Men in Tournaments. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift Für Die Gesamte Staatswissenschaft*, 169(4), 621–659. <https://www.jstor.org/stable/24548975>
- Davis, A. M., & Hyndman, K. (2019). Multidimensional Bargaining and Inventory Risk in Supply Chains: An Experimental Study. *Management Science*, 65(3), 1286–1304. <https://doi.org/10.1287/mnsc.2017.2985>
- Davis, A. M., & Hyndman, K. (2021). Private Information and Dynamic Bargaining in Supply Chains: An Experimental Study. *Manufacturing & Service Operations Management*, 23(6), 1449–1467. <https://doi.org/10.1287/msom.2020.0896>
- Eil, D., & Rao, J. M. (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics*, 3(2), 114–138. <https://doi.org/10.1257/mic.3.2.114>
- Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review*, 14(1), 57–74. <https://doi.org/10.2307/258191>
- Fershtman, C., & Gneezy, U. (2001). Strategic Delegation: An Experiment. *The RAND Journal of Economics*, 32(2), 352–368. <https://doi.org/10.2307/2696414>

- Gächter, S., & Riedl, A. (2005). Moral Property Rights in Bargaining with Infeasible Claims. *Management Science*, 51(2), 249–263. <https://doi.org/10.1287/mnsc.1040.0311>
- Galizzi, M. M., & Navarro-Martinez, D. (2019). On the External Validity of Social Preference Games: A Systematic Lab-Field Study. *Management Science*, 65(3), 976–1002. <https://doi.org/10.1287/mnsc.2017.2908>
- Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information Avoidance. *Journal of Economic Literature*, 55(1), 96–135. <https://doi.org/10.1257/jel.20151245>
- Grossman, S. J., & Hart, O. D. (1983). An Analysis of the Principal-Agent Problem. *Econometrica*, 51(1), 7–45. <https://doi.org/10.2307/1912246>
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367–388. [https://doi.org/10.1016/0167-2681\(82\)90011-7](https://doi.org/10.1016/0167-2681(82)90011-7)
- Hamman, J. R., Loewenstein, G., & Weber, R. A. (2010). Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship. *American Economic Review*, 100(4), 1826–1846. <https://doi.org/10.1257/aer.100.4.1826>
- Harsanyi, J. C., & Selten, R. (1972). A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information. *Management Science*, 18(5-part-2), 80–106. <https://doi.org/10.1287/mnsc.18.5.80>
- Haruvy, E., Katok, E., & Pavlov, V. (2020). Bargaining Process and Channel Efficiency. *Management Science*, 66(7), 2845–2860. <https://doi.org/10.1287/mnsc.2019.3360>
- Hetherington, M. J., & Rudolph, T. J. (2015). *Why Washington Won't Work: Polarization, Political Trust, and the Governing Crisis*. University of Chicago Press.
- Kennan, J., & Wilson, R. (1990). Can Strategic Bargaining Models Explain Collective Bargaining Data? *The American Economic Review*, 80(2), 405–409. <https://www.jstor.org/stable/2006609>
- Leider, S., & Lovejoy, W. S. (2016). Bargaining in Supply Chains. *Management Science*, 62(10), 3039–3058. <https://doi.org/10.1287/mnsc.2015.2273>
- Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *The Journal of Economic Perspectives*, 21(2), 153–174. <https://www.jstor.org/stable/30033722>
- Lo, B., Field, M. J., & Institute of Medicine (US) Committee on Conflict of Interest in Medical Research, E. (2009). Policies on Conflict of Interest: Overview and Evidence. In *Conflict of Interest in Medical Research, Education, and Practice*. National Academies Press (US).
- Melnikoff, D. E., & Strohminger, N. (2020). The automatic influence of advocacy on lawyers and novices. *Nature Human Behaviour*, 4(12), 1258–1264. <https://doi.org/10.1038/s41562-020-00943-3>
- Mnookin, R. H., & Susskind, L. E. (1999). *Negotiating on Behalf of Others: Advice to Lawyers, Business Executives, Sports Agents, Diplomats, Politicians, and Everybody Else*. SAGE Publications.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. <https://doi.org/10.1037/0033-295x.115.2.502>
- Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, 29(3), 315–335. <https://doi.org/10.2307/1909635>
- Nagarajan, M., & Bassok, Y. (2005). Bargaining and Alliances in Supply Chains. In G. J. Doukidis & A. P. Vrechopoulos (Eds.), *Consumer Driven Electronic Transformation* (pp.

- 39–51). Springer. https://doi.org/10.1007/3-540-27059-0_4
- Nash, J. F. (1950). The Bargaining Problem. *Econometrica*, 18(2), 155–162. <https://doi.org/10.2307/1907266>
- Neto, N. S. S., & Fontanari, J. F. (2023). On the efficacy of the wisdom of crowds to forecast economic indicators. *The European Physical Journal B*, 96(1), 6. <https://doi.org/10.1140/epjb/s10051-023-00482-6>
- Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, 17(1), 39–57. <https://doi.org/10.1002/bdm.460>
- Radzevick, J. R., & Moore, D. A. (2011). Competing to Be Certain (But Wrong): Market Dynamics and Excessive Confidence in Judgment. *Management Science*, 57(1), 93–106. <https://doi.org/10.1287/mnsc.1100.1255>
- Rangaswamy, A., & Shell, G. R. (1997). Using Computers to Realize Joint Gains in Negotiations: Toward an "Electronic Bargaining Table". *Management Science*, 43(8), 1147–1163. <https://www.jstor.org/stable/2634576>
- Ross, S. A. (1973). The Economic Theory of Agency: The Principal's Problem. *The American Economic Review*, 63(2), 134–139. <https://www.jstor.org/stable/1817064>
- Sadka, J., Seira, E., & Woodruff, C. (2024). Information and Bargaining through Agents: Experimental Evidence from Mexico's Labour Courts. *The Review of Economic Studies*, 91(6), 3677–3711. <https://doi.org/10.1093/restud/rdae003>
- Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes*, 121(2), 246–255. <https://doi.org/10.1016/j.obhdp.2013.02.001>
- Samuelson, W. (1980). First-Offer Bargains. *Management Science*, 26(2), 155–164. <https://www.jstor.org/stable/2630249>
- Schotter, A., Zheng, W., & Snyder, B. (2000). Bargaining Through Agents: An Experimental Study of Delegation and Commitment. *Games and Economic Behavior*, 30(2), 248–292. <https://doi.org/10.1006/game.1999.0728>
- Schwardmann, P., Tripodi, E., & van der Weele, J. J. (2022). Self-Persuasion: Evidence from Field Experiments at International Debating Competitions. *American Economic Review*, 112, 1118–1146.
- Schwardmann, P., & van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, 3(10), 1055–1061. <https://doi.org/10.1038/s41562-019-0666-7>
- Schweinsberg, M., Ku, G., Wang, C. S., & Pillutla, M. M. (2012). Starting high and ending with nothing: The role of anchors and power in negotiations. *Journal of Experimental Social Psychology*, 48(1), 226–231. <https://doi.org/10.1016/j.jesp.2011.07.005>
- Shachat, J., & Swarthout, J. T. (2013). Auctioning the Right to Play Ultimatum Games and the Impact on Equilibrium Selection. *Games*, 4(4), 738–753. <https://doi.org/10.3390/g4040738>
- Shachat, J., & Tan, L. (2015). An Experimental Investigation of Auctions and Bargaining in Procurement. *Management Science*, 61(5), 1036–1051. <https://doi.org/10.1287/mnsc.2013.1880>
- Shalvi, S., Soraperra, I., van der Weele, J., & Villeval, M. C. (2019). Shooting the Messenger? Supply and Demand in Markets for Willful Ignorance. In *Tinbergen Institute Discussion Papers* (19-071/I). Tinbergen Institute.

- Snizek, J. A., & Van Swol, L. M. (2001). Trust, Confidence, and Expertise in a Judge-Advisor System. *Organizational Behavior and Human Decision Processes*, 84(2), 288–307. <https://doi.org/10.1006/obhd.2000.2926>
- Tang, W., Bearden, J. N., & Tsetlin, I. (2009). Ultimatum Deadlines. *Management Science*, 55(8), 1423–1437. <https://doi.org/10.1287/mnsc.1090.1034>
- Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, 44(5), 1368–1375. <https://doi.org/10.1016/j.jesp.2008.04.006>
- U.S. Chamber Institute for Legal Reform. (2018). *Costs and Compensation of the U.S. Tort System*.
- Van Zant, A. B. (2022). Strategically overconfident (to a fault): How self-promotion motivates advisor confidence. *Journal of Applied Psychology*, 109–129. <https://doi.org/10.1037/apl0000879>
- Wang, X., & Navarro-Martinez, D. (2023). Bridging the gap between the economics lab and the field: Dictator games and donations. *Judgment and Decision Making*, 18, e18. <https://doi.org/10.1017/jdm.2023.19>

Appendix A
Supplementary Results of Study 1

Figure A1

The range covered by each of the five quintiles for Plaintiffs and Defendants in Study 1.

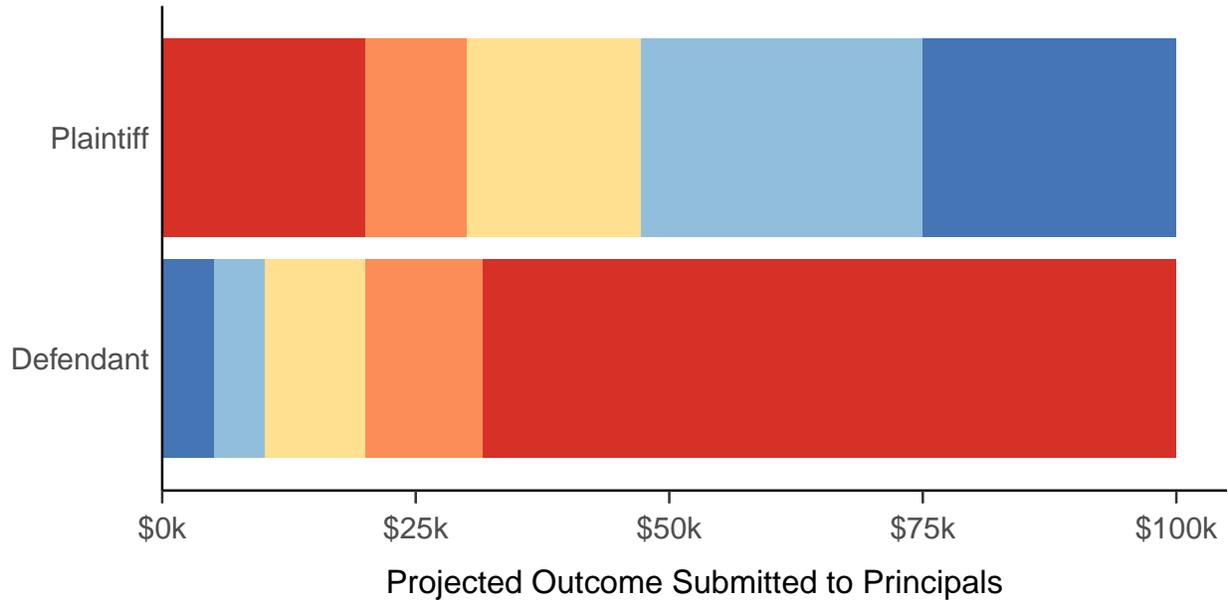
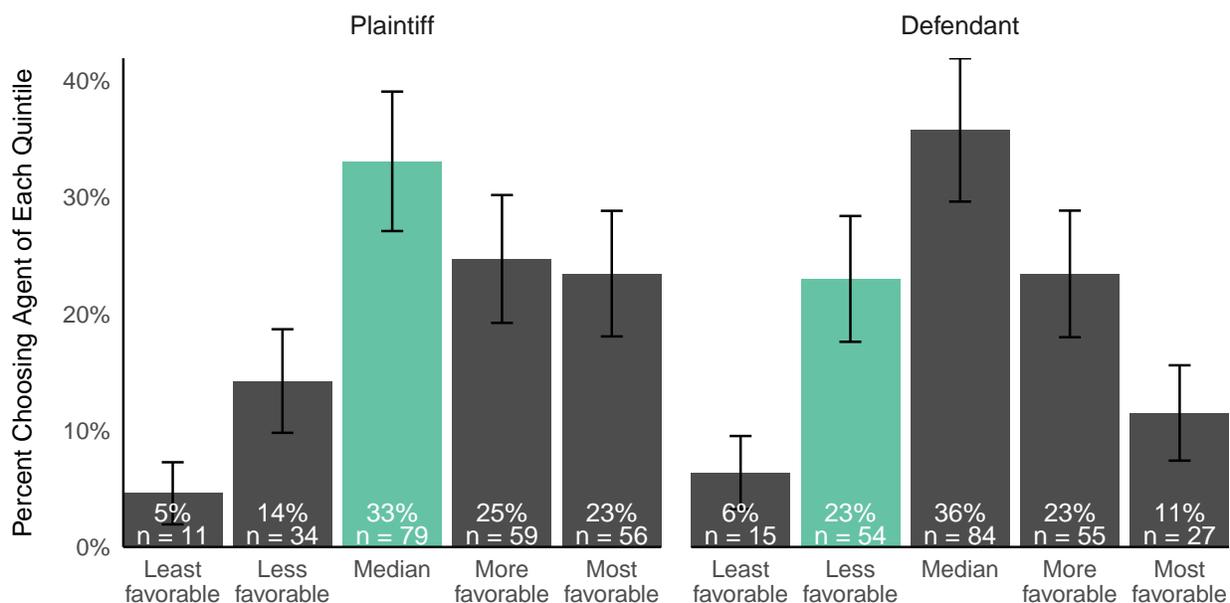


Figure A2

The share of principals choosing an Agent from each of the respective quintiles in Study 1. Notably, the median quintile is the most popular choice for both parties. However, Defendants are more likely than Plaintiffs to hire an Agent who made one of the two lowest forecasts, while Plaintiffs are more likely to hire an Agent who made the highest forecasts. Error bars show 95% confidence intervals.

***Alternative analysis for Study 1 following Preregistration***

We preregistered an analysis to see if Principals are more likely to choose an Agent who is more favorable to their position. This analysis was overly complicated, but we report it here for completeness. Specifically, we look at each set of Agents presented to Principals and for each Principal create five observations in which the predictor variable for each of the five quintiles recoded such that “1” is least favorable and “5” is most favorable. We then set an outcome measure that is “1” if the Agent from that quintile was selected and “0” if not. We conduct an OLS regression and cluster standard errors at the Agent level (Column 1). We correct this analysis by also clustering standard errors at the Principal level (Column 2), and at both levels (Column 3). Across all specifications, Principals are indeed more likely to choose an Agent who is more favorable to them.

Table A1

Preregistered analysis for Study 1, with Agent Quintile recoded such that a higher value is more favorable to the Principal. We find that Principals are responsive to how favorable the Agent is to them and select those from higher quintiles more frequently.

	(1)	(2)	(3)
Agent Quintile	0.030*** (0.006)	0.030*** (0.005)	0.030*** (0.006)
Constant	0.111*** (0.018)	0.111*** (0.015)	0.111*** (0.018)
N	2370	2370	2370
Clustered SE	Agent	Principal	Agent + Principal

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A2

Robustness check for Study 1, including participants who made inconsistent settlement offers. Column 1 tests whether belief polarization (Defendant x All Agents interaction) is greater among Selected Agents. Columns 2–3 test whether Selected Agents earn less than All Agents for each role. Column 4 tests whether impasse rates are higher among Selected Agents (negative coefficient on All Agents indicates lower impasse). All results are consistent with the main analysis.

	Beliefs	Plaintiff Earnings	Defendant Earnings	Impasse
Defendant	-16 367.751*** (1765.677)			
All Agents	-5035.012** (1752.753)	0.081*** (0.009)	0.089*** (0.016)	-0.142*** (0.015)
Defendant x All Agents	9179.936*** (2500.048)			
Constant	38 130.644*** (1243.243)	1.173*** (0.007)	1.882*** (0.012)	0.788*** (0.011)
N	946	481	465	946

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Appendix B
Supplementary Results of Study 2

Figure B1

The range covered by each of the five quintiles for Plaintiffs and Defendants in Study 2.

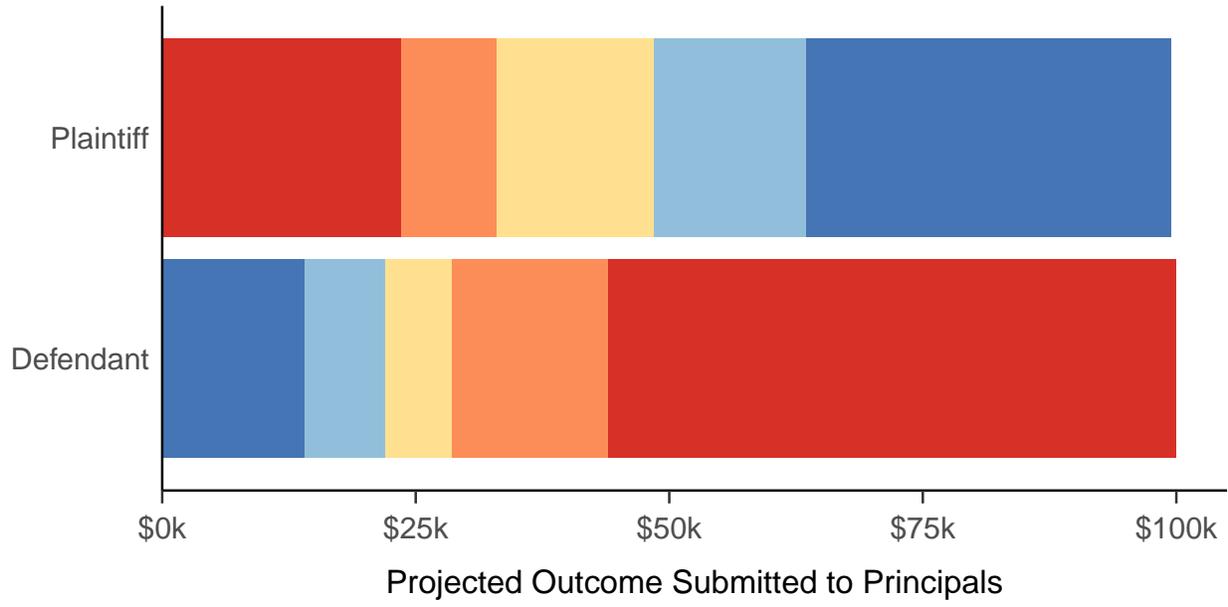


Figure B2

The share of principals choosing an Agent from each of the respective quintiles in Study 2. Error bars show 95% confidence intervals.

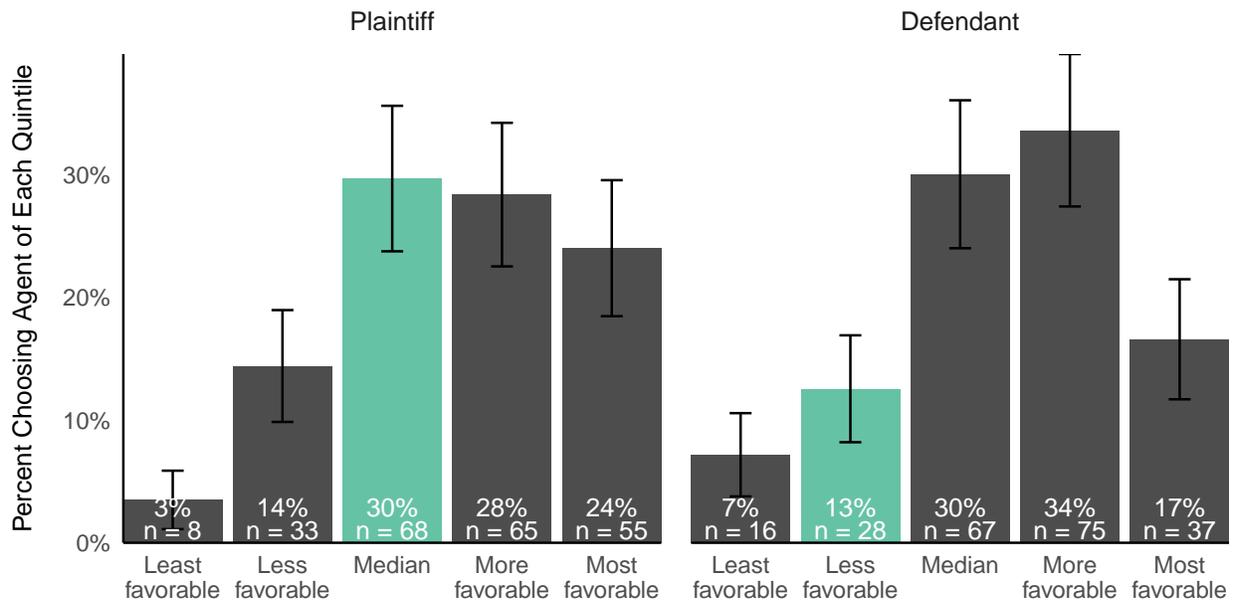


Figure B3

Expected payoffs for the Plaintiff and the Defendant when choosing an agent from each quintile, given each quintile the other party could have selected from (Study 2). The first number in each cell shows the earnings for the Defendant (row player), the second shows earnings for the Plaintiff (column player). Cell shading indicates the impasse rate (blue = low impasse, orange = high impasse). The purple border marks the Nash Equilibrium (Plaintiff quintile 3, Defendant quintile 4).

Defendant Quintile Selected	(5) Least favorable	(\$2.56, \$1.40) Impasse: 3%	(\$2.04, \$1.73) Impasse: 19%	(\$1.89, \$1.81) Impasse: 26%	(\$1.67, \$1.75) Impasse: 49%	(\$1.60, \$1.54) Impasse: 71%
	(4)	(\$2.77, \$1.19) Impasse: 4%	(\$2.22, \$1.43) Impasse: 29%	(\$2.03, \$1.45) Impasse: 43%	(\$1.83, \$1.33) Impasse: 70%	(\$1.75, \$1.22) Impasse: 86%
	(3)	(\$2.89, \$1.02) Impasse: 7%	(\$2.19, \$1.21) Impasse: 50%	(\$1.96, \$1.20) Impasse: 70%	(\$1.81, \$1.15) Impasse: 87%	(\$1.77, \$1.12) Impasse: 92%
	(2)	(\$2.85, \$0.91) Impasse: 20%	(\$2.04, \$1.12) Impasse: 70%	(\$1.84, \$1.15) Impasse: 85%	(\$1.78, \$1.11) Impasse: 93%	(\$1.76, \$1.08) Impasse: 96%
	(1) Most favorable	(\$2.45, \$0.86) Impasse: 58%	(\$1.79, \$1.05) Impasse: 96%	(\$1.73, \$1.07) Impasse: 100%	(\$1.73, \$1.07) Impasse: 100%	(\$1.74, \$1.07) Impasse: 100%
		(1) Least favorable	(2)	(3)	(4)	(5) Most favorable
		Plaintiff Quintile Selected				

Table B1

Robustness of the ex post optimal pairing to varying levels of value destruction in Study 2. The 30% row (bold) reflects the actual experimental parameterization. t-tests compare per-agent mean earnings under random pairing (All) vs. observed participant selection (Selected).

Destruction	Nash Eq.	Plaintiff			Defendant		
		All	Selected	t-test	All	Selected	t-test
0%	(mixed only)	\$1.54	\$1.54	.772	\$2.46	\$2.46	.817
10%	P4, D3	\$1.45	\$1.43	.266	\$2.31	\$2.27	.033
20%	P3, D4	\$1.35	\$1.31	.002	\$2.16	\$2.08	< .001
30%	P3, D4	\$1.26	\$1.19	< .001	\$2.01	\$1.89	< .001
40%	P2, D4	\$1.17	\$1.08	< .001	\$1.86	\$1.70	< .001
50%	P2, D4	\$1.08	\$0.96	< .001	\$1.72	\$1.51	< .001

Table B2

Robustness check for Study 2: Beliefs about the fair outcome (Judge's Ruling), including participants who made inconsistent settlement offers. All four columns replicate the corresponding specifications from the main analysis. Results are consistent with the main analysis.

	Self vs. Selected	Self vs. All	All vs. Selected	Self vs. Principals
Defendant	-6343.809*** (1616.695)	-6343.809*** (1736.774)	-4884.948** (1658.725)	-9840.141*** (1628.463)
Agent	5490.216*** (1642.467)	597.872 (1753.015)	4892.344** (1658.725)	
Defendant x Agent	-7996.098*** (2320.139)	1458.860 (2475.824)	-9454.958*** (2361.316)	
Self-Representation				-481.483 (1607.260)
Defendant x Self-Representation				3496.333 (2270.406)
Constant	36 376.596*** (1153.866)	36 376.596*** (1239.569)	36 974.468*** (1165.288)	36 858.079*** (1143.829)
N	931	943	916	931

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B3

Robustness check for Study 2: Earnings and impasse rates, including participants who made inconsistent settlement offers. Columns 1–2 test whether Selected Agents earn less than Self-Representation for each role. Columns 4–5 test whether Selected Agents earn less than All Agents. Columns 3 and 6 test whether impasse rates differ. All results are consistent with the main analysis.

	Self vs. Selected			All vs. Selected		
	Plaintiff \$	Defendant \$	Impasse	Plaintiff \$	Defendant \$	Impasse
Self-Representation	0.109*** (0.012)	0.068*** (0.017)	-0.148*** (0.016)			
All Agents				0.100*** (0.012)	0.080*** (0.020)	-0.150*** (0.017)
Constant	1.207*** (0.008)	1.917*** (0.012)	0.730*** (0.012)	1.207*** (0.009)	1.917*** (0.014)	0.730*** (0.012)
N	464	467	931	464	452	916

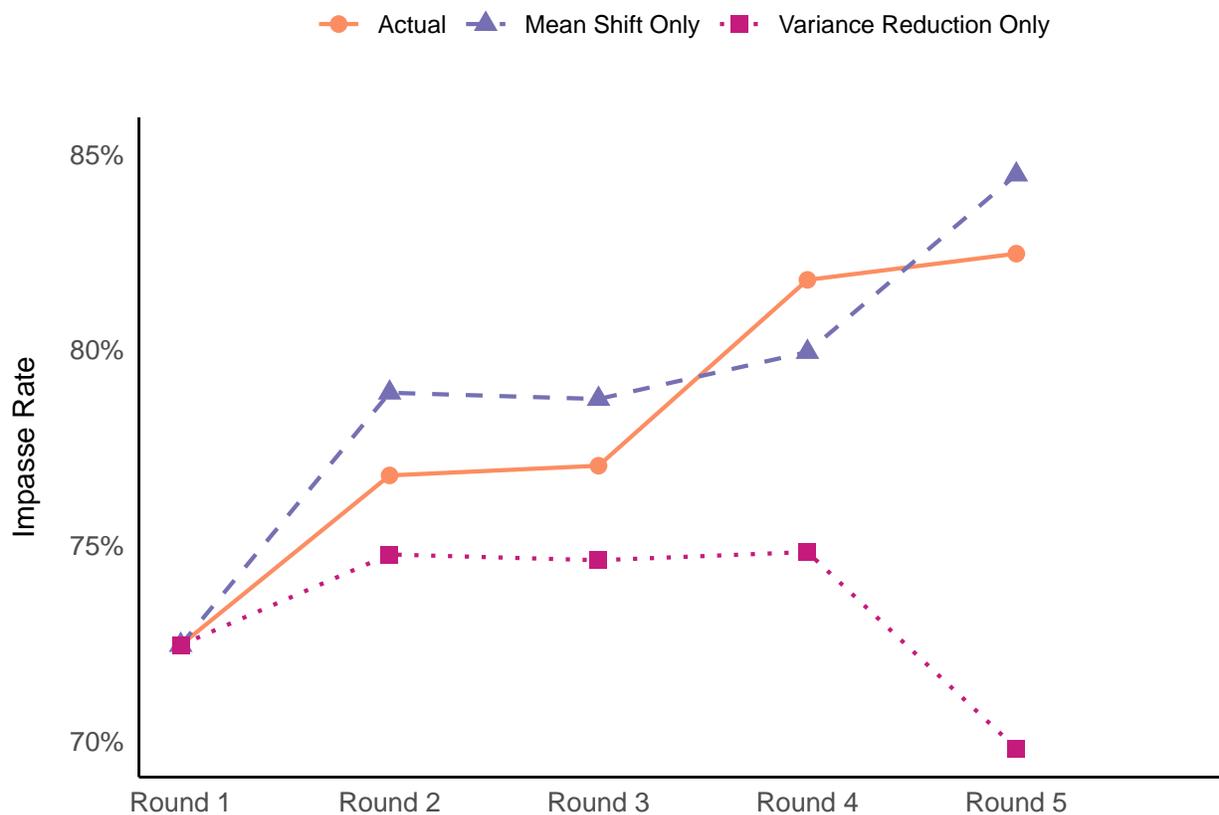
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Appendix C

Supplementary Results of Study 3

Figure C1

Decomposition of the rising impasse rate in Study 3 into mean shift and variance reduction channels. The solid orange line shows the actual impasse rate among Selected Agents across five rounds. The dashed purple line shows the counterfactual impasse rate when only the mean shift in proposals operates (with variance held at Round 1 levels). The dotted pink line shows the counterfactual when only variance reduction operates (with the mean held at Round 1 levels). The mean shift channel is the dominant driver of the escalating impasse rate.



Appendix D

Supplementary Results of Study 4

Figure D1

Payoff matrix showing average bonus earnings for Buyers and Sellers across all available pairings of beliefs about the outcomes in Study 4. Each cell reports (Buyer Bonus, Seller Bonus), the number of pairs observed, and the impasse rate. Empty cells indicate belief pairings with no observations.

Buyer's Expected Outcome	\$4.0M (Least favorable)	--	(\$1.50, \$1.00) n = 1 Impasse: 0%	(\$1.50, \$1.00) n = 1 Impasse: 0%	--	(\$0.00, \$2.50) n = 1 Impasse: 0%	(\$0.20, \$2.30) n = 1 Impasse: 0%
	\$3.5M	--	(\$1.10, \$1.40) n = 4 Impasse: 0%	(\$1.00, \$1.50) n = 4 Impasse: 0%	(\$1.00, \$1.50) n = 1 Impasse: 0%	(\$0.86, \$1.43) n = 7 Impasse: 43%	(\$0.80, \$1.40) n = 5 Impasse: 60%
	\$3.0M	(\$2.00, \$0.50) n = 1 Impasse: 0%	(\$1.33, \$1.06) n = 9 Impasse: 22%	(\$1.27, \$1.17) n = 15 Impasse: 13%	(\$1.07, \$1.13) n = 15 Impasse: 27%	(\$0.90, \$1.38) n = 22 Impasse: 45%	(\$0.90, \$1.27) n = 15 Impasse: 67%
	\$2.5M	(\$1.00, \$1.00) n = 1 Impasse: 100%	(\$1.34, \$1.01) n = 7 Impasse: 29%	(\$1.19, \$1.04) n = 18 Impasse: 28%	(\$1.02, \$1.27) n = 20 Impasse: 40%	(\$1.05, \$1.19) n = 19 Impasse: 53%	(\$1.08, \$1.10) n = 23 Impasse: 65%
	\$2.0M	(\$1.90, \$0.60) n = 1 Impasse: 0%	(\$1.40, \$1.10) n = 2 Impasse: 0%	(\$1.24, \$0.81) n = 10 Impasse: 70%	(\$1.12, \$1.05) n = 23 Impasse: 65%	(\$1.05, \$1.03) n = 24 Impasse: 83%	(\$1.07, \$1.16) n = 18 Impasse: 56%
	\$1.5M (Most favorable)	--	(\$1.00, \$1.00) n = 2 Impasse: 100%	(\$0.95, \$1.30) n = 4 Impasse: 50%	(\$1.20, \$0.90) n = 5 Impasse: 80%	(\$1.12, \$0.94) n = 8 Impasse: 88%	(\$0.98, \$1.11) n = 11 Impasse: 82%
		\$1.5M (Least favorable)	\$2.0M	\$2.5M	\$3.0M	\$3.5M	\$4.0M (Most favorable)
		Seller's Expected Outcome					