

# The Winner's Curse: Conditional Reasoning and Belief Formation

Christian Koch\* and Stefan P. Penczynski†

June 2017

## Abstract

In explaining the winner's curse, recent approaches have focused on one of two cognitive processes: conditional reasoning and belief formation. We provide the first joint experimental analysis of the role of these two obstacles. First, we observe that overbidding decreases significantly between a simple common-value auction and a transformed version of this auction that does not require conditional reasoning. Second, assistance in belief formation leads to comparable behavioral changes in both games. The two effects are of similar magnitude and amplify each other when jointly present. We conclude that the combination and the interaction of the two cognitive processes in auctions lead to relatively low strategic sophistication compared to other domains. The study's focus on games' objective cognitive challenges is potentially useful for predictions across games and complements the common focus on behavioral models and their explanatory power.

*JEL classification:* D44, D82, C91

*Keywords:* Auctions, Winner's curse, Conditional Reasoning, Beliefs

---

\**Corresponding Author.* New York University Abu Dhabi (NYUAD), A5 - 1131, PO Box 129188, Abu Dhabi, UAE. Email: christian.koch@nyu.edu.

†University of Mannheim, L7, 3-5, 68131 Mannheim, Germany. Email: stefan.penczynski@uni-mannheim.de.

# 1 Introduction

The “winner’s curse” (WC) in common-value auctions (CVA) refers to the systematic overbidding relative to the Bayesian Nash equilibrium (BNE) that leads to losses for winners in field settings and laboratory experiments.<sup>1</sup> This phenomenon is one of the most important and robust findings in empirical auction analysis and has generated ample theoretical work.

Two main departures from the BNE have been modeled. Both maintain the assumption that players best respond to their beliefs but relax the requirement of *consistency of beliefs*. First, in equilibrium models such as cursed equilibrium (CE, Eyster and Rabin, 2005), behavioral equilibrium (Esponda, 2008), and the application of analogy-based expectation equilibrium to auctions (Jehiel, 2005; Jehiel and Koessler, 2008), beliefs do not fully take into account what bids tell about underlying signals, capturing that agents do not optimally adjust for the information revealed by winning. Second, the level- $k$  model assumes non-equilibrium beliefs that result from iterated best responses (Nagel, 1995; Stahl and Wilson, 1995). It has been applied to private information games such as auctions and zero-sum betting (Crawford and Iriberri, 2007; Brocas, Carrillo, Wang, and Camerer, 2014). When one assumes beliefs of uninformed play, this approach can implicitly capture that agents do not fully account for revealed information.

Doubts have been cast on the sufficiency of these *belief-based* models to explain auction behavior. With an innovative semi-computerized version of the maximal game, Ivanov, Levin, and Niederle (2010, ILN) experimentally study whether these models can explain the WC and claim that they cannot. Along these lines, Charness and Levin (2009, CL) use computerized sellers in an acquiring-a-company game and document that subjects have a more general problem with *conditional reasoning* – drawing appropriate conclusions from hypothetical events – that seems not to be fully captured by the relaxation of beliefs.

In turn, however, Costa-Gomes and Shimoji (2015) criticize ILN’s use of game theoretical concepts when the interaction with a known computer program is a single-person decision problem. They argue that belief-based models are indeed compatible with some observations from ILN’s experiment. Moreover, Camerer, Nunnari, and Palfrey (2015) suggest on the basis of the Quantal Response

---

<sup>1</sup>See Capen, Clapp, and Campbell (1971) and Roll (1986) for evidence from the oil industry and corporate takeovers, respectively, and Bazerman and Samuelson (1983), Kagel and Levin (1986), Avery and Kagel (1997), Goeree and Offerman (2002), Lind and Plott (1991), Grosskopf, Bereby-Meyer, and Bazerman (2007), and the literature discussed in Kagel and Levin (2002) for experimental evidence.

Equilibrium (QRE, McKelvey and Palfrey, 1995) that imprecise best responses combined with non-equilibrium beliefs could explain observed behavior.

This discussion shows that no consensus has been reached on how to explain the WC. In this study, we do not test concrete models of reasoning, but take a step back and focus on two objective game complexities whose relative importance in causing the WC is – as shown above – disputed in the literature: the needs for *conditional reasoning* and for *belief formation*. Both activities are indispensable to reach a best response. In any strategic situation, subjects have to *form beliefs* about their opponents’ behavior in order to know what to best respond to. In CVAs, best responding further requires *conditioning* on hypothetical situations induced by the game’s structure. For example, one’s bid is only relevant when winning, which implies that all others have bid less. Crucially, which of the two complexities poses a more substantial challenge for bidders in CVAs remains an open empirical question. By providing the first joint experimental analysis that disentangles the impact of these two cognitive processes in a CVA setting, we are able to determine – as our paper’s main contribution – whether the WC is predominantly driven by *conditional reasoning* or *belief formation*. Studying strategic behavior with a focus on objective game complexities enables us to establish how two of these complexities – which can be found in a variety of important game – generally affect behavior. Notably, this analysis is not constrained by a more specific structure on how people think about these problems. Physics can predict the bending of a horizontal steel bar due to vertical forces without a detailed model of the tensions inside the bar. Similarly, we propose to relate deviations from equilibrium play to objective game complexities such as the need for conditional reasoning or for elements of belief formation. This approach has potential for improving predictions across very different games, an area of study so far put in second place.

Our starting point is a simple first-price CVA adapted from Kagel and Levin (1986). At the core of our investigation is a transformation of this game that maintains the strategic nature of the original auction game in terms of best response functions and equilibria but removes the need to engage in conditional reasoning. This allows us to cleanly identify the effect that this cognitive activity has on bidding behavior and the WC. Independently of this variation, we further change the need to form beliefs in two ways. First, we *fully* remove the need to form beliefs by letting subjects play against naïve computer opponents that follow a known simple strategy. Second, we *partially* remove crucial parts of belief formation but maintain the strategic uncertainty associated with human opponents when we let subjects play against human opponents after they played the computer.

The preceding encounter with the naïve computer provides subjects with a first scenario or basic belief about how opponents could behave.

Following our focus on objective cognitive complexities, we provide a simple formalization to measure the complexities' behavioral impact in a flexible and general way. Defining a measure  $\mu$  on the action space, we normalize the distance to 1 between equilibrium play,  $\mu^e = 0$ , and uninformed random play,  $\mu^u = 1$ . We judge a cognitive complexity by the sign and magnitude of the change  $\Delta\mu$  caused in the direction away from optimal behavior.

In the modified auction setting that requires neither conditional reasoning nor belief formation for optimal behavior, we observe bids that are close to equilibrium play with  $\mu$  of 0.29. From there, we obtain three main results. First, introducing the need to condition – without requiring any belief formation – increases bids significantly and moves them further away from optimal play with  $\Delta\mu = 0.18$ . In addition, it increases the incidence of the WC by 15 percentage points. Second, requiring partial or full belief formation – in the absence of conditional reasoning – leads to remarkably similar increases in bids,  $\Delta\mu$  of 0.15 and 0.20, and raises the number of subjects falling prey to the WC by 8 to 11 percentage points, respectively. Interestingly, the partial belief manipulation suggests that the mere need to form a first belief, at a given level of strategic uncertainty, already proves challenging for subjects. Third, no generally significant differences emerge when comparing the magnitude of the effects of conditional reasoning and belief formation. Although both effects individually worsen game play, the fraction of plausible bids still remains non-negligible. Combining conditional reasoning and full belief formation results in behavior fairly far away from equilibrium,  $\mu = 0.81$ , as usually observed in CVA settings. Interestingly, the combination of both effects,  $\Delta\mu = 0.52$ , leads further away from equilibrium than expected by the sum of the two individual effects,  $\Delta\mu = 0.38$ , implying that the two strengthen each other and exhibit what we call *cognitive diseconomies*.

The two cognitive complexities jointly produce an extreme case of game-dependent sophistication that is not fully captured by belief-based models. This explains why CL and ILN do not find support for those models when at the same time such support is abundant in other domains in which conditional reasoning is not required (see Crawford, Costa-Gomes, and Iriberri, 2013).

A number of further papers are closely related to our work. Levin et al. (2016) analyze the conditioning problem in the WC in more detail by separating the involved Bayesian updating from non-probabilistic reasoning. In particular, the authors compare results from a first-price auction with a strategically equivalent

Dutch-CVA that makes the conditioning problem more salient. Relatedly, but in non-auction settings, Esponda and Vespa (2014), Louis (2015), and Ngangoue and Weizsäcker (2015) have analyzed conditioning in more depth by separating two involved steps – hypothetical thinking per se and conditioning on hypothetical events – and comparing behavior in simultaneous and sequential games. With our transformation, we propose a complementary way of studying conditional reasoning in auction settings. Crucially, we do not provide a differentiated analysis of conditional reasoning itself but relate the impact of the overall conditioning effect in causing the WC to the impact of belief formation. Moreover, Charness, Levin, and Schmeidler (2014) observe the WC in a generalized information environment in which bidders hold identical and public information. Their innovative design allows them to disentangle the influence of heterogeneity in estimating the common value from non-optimal bidding behavior. They show that both are relevant for the WC. Complementing their results, our study only focuses on the bidding behavior but additionally disentangles the role of conditional reasoning and belief formation. Finally, Levin and Reiss (2012) construct a behavioral auction design in which the payment rule incorporates the adverse selection problem that is at the origin of the WC. They observe that the WC is still present in their data. The authors adjust the payment rule but do not transform the auction game as we do.

Due to our method of transformation, our paper also relates to the broad set of studies that investigate behavior using strategically very similar games. The largest fraction of those studies considers framing effects that influence subjects' behavior but do not result from the strategic nature of the situation (for example Tversky and Kahneman, 1986; Osborne and Rubinstein, 1994; Chou et al., 2009). Another methodologically interesting instance of strategic equivalence is the experimental, so-called “strategy method” in which participants make contingent decisions for all decision nodes that they will possibly encounter in a game (Brandts and Charness, 2011). In a different manner, strategically equivalent versions of a game can facilitate the investigation of particular aspects of behavior. For example, Nagel and Tang (1998) use a repeated, normal-form centipede game to investigate learning behavior without aspects of sequential reciprocity.

In our study, we craft two similar games that differ in the cognitive process under investigation: conditional reasoning. To the best of our knowledge, our experiment is the first that uses such a transformation as a means to investigate the impact of this particular cognitive activity in strategic reasoning. By this virtue, our approach opens further avenues for investigation in settings with similar cognitive processes. For example, conditioning on being pivotal in a jury decision

is part of strategic voting (Feddersen and Pesendorfer, 1998) and conditioning on message selection is part of being optimally persuaded (Glazer and Rubinstein, 2004).

## 2 Design and Hypotheses

In our experimental design, we will use two different games: a simplified standard *auction game* that serves as the basis for constructing a *transformed game* which does *not* require conditional reasoning. The starting point for both games is a standard CVA setting as in Kagel and Levin (1986). At the beginning of each period, the common value of the auctioned item  $W^* \in [\underline{W}, \overline{W}]$  is randomly chosen, with all values equally likely. Each of the  $n$  bidders receives a private signal  $x_i \in [W^* - \delta, W^* + \delta]$ , with  $\delta > 0$ , which is informative about this common value. Bidders make bids  $a_i$  in a sealed-bid first-price auction in which the highest bidder wins the auction and pays his bid. The payoff of the highest-bidding player is  $u_i = W^* - a_i$ . In case a bidder does not make the highest bid, his payoff is  $u_i = 0$ .

### 2.1 The Games

#### 2.1.1 Auction Game

We simplify this general setting – to be able to later construct the transformed game – mainly by allowing only for two signals and two players  $n = 2$ . The common value  $W^*$  is uniformly distributed in the interval  $[25, 225]$ . Bidders receive a private binary signal  $x_i \in \{W^* - 3, W^* + 3\}$  drawn *without replacement*. Thus, when receiving signal  $x_i$ , the set  $\Omega_i$  of possible item values  $W^*|x_i$  has only two elements  $\{x_i - 3, x_i + 3\}$  that are equally likely (when disregarding boundary signals). Let us denote the state of  $i$ 's world by  $\omega_i = \{h, l\}$ , indicating whether  $W^*$  is high or low relative to  $i$ 's signal, so that  $W^*|h = x_i + 3$  and  $W^*|l = x_i - 3$ .<sup>2</sup> Notably,  $W^*|h$  implies that  $i$  has received the *lower* of the two possible signals and, thus,  $x_j|h = x_i + 6$ , while  $W^*|l$  reflects that  $i$  received the *higher* signal, implying  $x_j|h = x_i - 6$ . Moreover, it thus follows that potential item values differ for the two players,  $\Omega_i \neq \Omega_j$ . To ensure an equilibrium in pure strategies, we only allow bids  $a_i \in [x_i - 8, x_i + 8]$ . As a tie-breaker in case of identical bids, the lower-signal player wins the auction. As in Kagel and Levin (1986), bids are made in a first-price sealed bid auction.

---

<sup>2</sup>In the remainder, for ease of exposition, any state indication  $\omega$  will refer to  $\omega_i$  and not  $\omega_j$ .

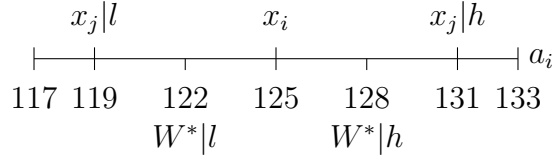


Figure 1: The example in the auction game.

Consider the following example as illustrated in figure 1. The male player  $i$  receives the signal  $x_i = 125$ . Hence,  $i$ 's bids are limited to  $a_i \in [117, 133]$ . Moreover,  $i$  knows that with 0.5 probability the value of the item is  $W^*|h = 128$  or  $W^*|l = 122$ ,  $\Omega_i = \{122, 128\}$ . When subjects make their decisions in the experiment, the computer presents these two values to the subjects as shown in figure 2. Player  $i$  now has to understand that his female opponent  $j$  either has received the signal  $x_j|l = 119$  or  $x_j|h = 131$ .<sup>3</sup> When finding the best response to his beliefs,  $i$  has to condition on these two hypothetical events.

Figure 2: Screenshot auction game: “**Trading period 1:** Your private information signal is 125.00 Taler. Hence, the true commodity’s value is either 122.00 or 128.00 Taler. **How much do you want to bid?**”

To further illustrate our game, let us now consider the example of symmetric bidding: player  $i$  thinks player  $j$  makes the same bid relative to her signal as he does, i.e.  $i$  thinks that  $j$  either bids  $a_j = a_i + 6$  or  $a_j = a_i - 6$ , depending on the state of the world. Due to the construction of the game,  $i$  knows that signals are always exactly 6 points apart but he does not know  $\omega$ . In this case, if  $i$  wins the auction,  $i$  has to *infer* that  $j$ 's bid must have been  $a_j = a_i - 6$  and that, for this reason,  $x_i > x_j$ . In a next step,  $i$  has to *infer* that the item value is lower than  $i$ 's signal,  $w = l$ . Put differently, for the case of symmetric bidding, the conditional value of the item  $E[W^*|win, x_i] = x_i - 3$ , is lower than the unconditional one,  $E[W^*|x_i] = x_i$  and player  $i$  has to take this into account

<sup>3</sup>If  $x_j = 119$ , then  $\Omega_j = \{116, 122\}$ . If  $x_j = 131$ , then  $\Omega_j = \{128, 134\}$ .

when deriving the best response. As outlined in more detail below, both players bid  $-8$  in equilibrium. Due to this symmetric bidding the discussed difference in expectations then matters.

### 2.1.2 Transformed Game

Let us now consider a different game, our transformed game. It is constructed as a CVA without private signals but with special auction rules and has the following structure: We again have two players. These players do not receive any signals but are instead informed about the two possible values an item can take,  $W_l^*$  or  $W_h^* = W_l^* + 6$ . Similar to the auction game, subjects are allowed to underbid  $W_l^*$  by 5 units and overbid  $W_h^*$  by 5 units,  $a_i \in [W_l^* - 5, W_h^* + 5]$ . In analogy to the auction game, the ranges of the values are  $W_l^* \in [25, 219]$  and  $W_h^* \in [31, 225]$ .<sup>4</sup>

Subjects are told that the realization of the two possible values depends on chance and on both players' bids. More precisely, instead of a simple first-price auction rule, subjects are provided with three special auction rules. First, if  $i$  overbids  $j$  by at least 6 units, he wins the auction for sure and either value realizes with probability of 0.5 ("rule 1"). Second, if  $i$  underbids  $j$  by at least 6 units, he does not win the auction and his payoff is 0 for sure ("rule 2"). Third, if the difference between both players' bids is smaller than 6 units, then the winning probability of each individual player is 0.5, and the smaller value  $W_l^*$  realizes irrespective of which player wins the auction. The loser obtains a payoff of 0 ("rule 3").

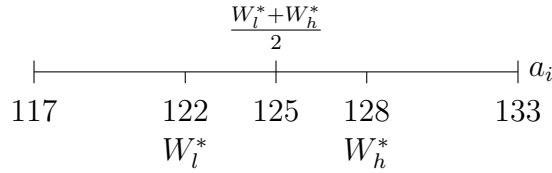


Figure 3: The example in the transformed game.

Before analyzing how the two presented games relate to each other, we illustrate the transformed game by considering an example as before (see figure 3). Player  $i$  as well as his opponent  $j$  are informed that the auctioned item either has value  $W_l^* = 122$  or  $W_h^* = 128$ . In the experiment, the decision screen presents these two values as shown in figure 4. Players' bids are limited to  $a_i \in [117, 133]$ . Notably,

<sup>4</sup>We choose the intervals such that the lowest and highest realizations are the same across the two games. Other ways of drawing this analogy are conceivable, however, this way is a straightforward one.



since there is no signal structure, player  $i$  does not have to think about hypothetical situations induced by the game structure. This has profound implications as can be seen by again considering the case of symmetric bidding. In the absence of individually varying signals, symmetric play simply implies that player  $i$  thinks that player  $j$  will bid  $a_j = a_i$ . Notably, in this case, the third rule is applicable and directly reveals – without requiring any further *inferences* as in the auction game – that the item value is low,  $W_l^*$ , completely independent of whether  $i$  or  $j$  wins. Put differently, the conditional and unconditional value of the item are the same for the case of symmetric bidding,  $E[W^*|win] = E[W^*] = W_l^*$ . As shown below, this matters in equilibrium where players symmetrically bid  $-8$ , just like in the auction game.

The screenshot shows a game interface with a light gray background. At the top, a text box contains the following German text: "Handelsperiode 1 : Der wahre Wert der Ware ist entweder 122.00 Taler oder 128.00 Taler. Dies hängt von Ihrem Gebot, vom Gebot des anderen Teilnehmers und vom Zufall ab." Below this text box is a larger box with the question "Wieviel möchten Sie bieten?". Underneath that is a box labeled "Gebot:" followed by a blue rectangular input field. At the bottom right of the interface is a red button with the text "Weiter".

Figure 4: Screenshot transformed game: “**Trading period 1:** The true commodity’s value is either 122.00 Taler or 128.00 Taler. The value depends on your bid, the other participant’s bid and chance. **How much do you want to bid?**”

### 2.1.3 Connection Between the Two Games

To see how the transformed game and the auction game relate to each other, we will analyze the underlying structure of the auction game in more detail. For this purpose, we express subjects’ strategies by *relative* bids. This is done with respect to their signal for the auction game,  $b_i = a_i - x_i$ , and with respect to the mean of the two potential item values for the transformed game,  $b_i = a_i - \frac{W_l^* + W_h^*}{2}$ . Due to their relevance, we will call these relative bids  $b_i$  just “bids” in the remainder and always specify when we talk about *absolute* bids  $a_i$ . Their use is only for analytical purposes, the instructions exclusively use absolute bids for both games and subjects make all their decisions in the same absolute metric.

Relative to the other player’s bid  $b_j$ ,  $i$ ’s strategies  $b_i$  lead to three kinds of interactions in the auction game, as illustrated in figure 5(i). First, if player  $i$  overbids player  $j$  – in relative terms – by at least 6 units,  $b_i \geq b_j + 6$ , he always

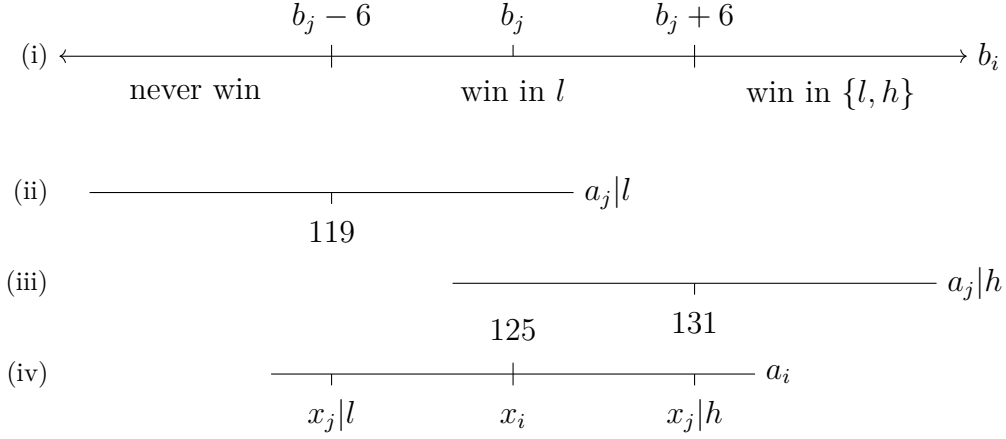


Figure 5: Three sets of relative bids  $b_i$  (i) induced by the relative position of signals  $x_i$ ,  $x_j|l$ , and  $x_j|h$ , illustrated for the case of  $b_j = 0$  (ii-iv).

wins the auction (“win in  $\{l, h\}$ ”). Such a bid bridges the distance even to  $x_j|h$  implying that  $i$  wins both in  $l$  and  $h$ . Second, conversely, if  $i$  underbids player  $j$  by at least 6 units,  $b_i \leq b_j - 6$ , he never wins the auction (“never win”). Third, if  $i$  bids less than 6 units away from  $j$ ’s bid,  $b_i \in (b_j - 6, b_j + 6)$ , he only wins the auction in  $l$ , i.e. he only wins with the higher signal implying a low item value  $W^*|l$  but not with the lower signal when the item value is high  $W^*|h$  (“win in  $l$ ”). This asymmetry reflects the standard adverse selection problem common to CVAs.

Figures 5(ii)-(iv) show how the three kinds of interaction derive from the position of signals. In particular, if player  $i$  has received a signal of 125 and *believes* that his opponent  $j$  just bids her signal,  $b_j = 0$ , player  $i$  has to *infer* that player  $j$ ’s absolute bid is either 119 or 131. Based on this inference, in case player  $i$  makes an absolute bid of at least 131, or  $b_i = 6$ , he will always win the auction. Bidding weakly below 119, or  $b_i = -6$ , results in never winning the auction. Finally, bidding above 119 but below 131 leads to winning only with the higher signal and, hence, with 50% chance. This results in an optimal bid of  $119 + \epsilon$ , with a small  $\epsilon > 0$ , for this set of strategies. Thus, within the “win in  $l$ ” set, the winner always receives the higher signal and the smaller item value realizes.

The transformed game is constructed such that  $W_l^*(= W^*|l)$  and  $W_h^*(= W^*|h)$  correspond to the possible item values from the perspective of a signal  $\frac{W_l^* + W_h^*}{2}$  in the auction game. Since relative bids only differ by a constant from absolute bids in the transformed game, the special rules of the transformed game exactly reflect the underlying structure of the auction game, as expressed in relative bids. Rule 1 captures overbidding by six points,  $b_i \geq b_j + 6$  (“win in  $\{l, h\}$ ”). Rule 2 reflects underbidding by six points,  $b_i \leq b_j - 6$  (“never win”), and rule 3 reflects

$b_i \in (b_j - 6, b_j + 6)$  (“win in  $l$ ”).

Let us reconsider the previous example from the perspective of the transformed game: Player  $i$  believes that his opponents  $j$  just bids the mean value of the item,  $a_j = 125$  or  $b_j = 0$ . Just by consulting the auction rules,  $i$  knows that bidding at least 131, or  $b_i = 6$ , will result in always winning the auction, bidding weakly below 119, or  $b_i = -6$ , will result in never winning the auction, and bidding in between will result in winning the item with the lower value with 50% chance.

The analysis of the underlying structure of the two games provides an intuition why the equilibrium of the two games is that both players bid  $-8$ . In this case, each player wins the auction with 50% chance (“win in  $l$ ” – “rule 3”) and the low item value realizes from the perspective of the winner,  $W^*|l = x - 3 = W_l^*$ . Deviating from this strategy is not profitable for two reasons. First, always winning the auction (“win in  $\{l, h\}$ ” – “rule 1”) with a six point higher bid of  $b = -2$  leads to losses in state  $l$  that are not sufficiently compensated in state  $h$ . Second, “never win” (“rule 2”) cannot result in positive profits. Notably, we ensure this equilibrium in pure strategies by restricting the action set to  $b \in [-8, 8]$ .<sup>5</sup>

The equilibria of the two games coincide because the special rules of the transformed game exactly reflect the underlying structure of the auction game, and we, thus, maintain the strategic nature of the latter game. Crucially, we remove the need to engage in what we call *conditional reasoning* in the transformed game. In the auction game, subjects first have to be able to think in hypothetical situations that are induced by the game’s structure. In our design, they have to realize that due to the signal structure two different hypothetical situations are possible: one’s own signal could be derived from a high or low item value. Afterwards, subjects have to condition on these two hypothetical events jointly when drawing appropriate conclusions about how to behave. In particular, they have to condition on the event of winning the auction. Crucially, the rules of the transformed game already reflect these two hypothetical events and present the conditional inferences from these events in form of winning probabilities and the realized item value. In other words and as illustrated in the example of symmetric bidding above, the rules spell out consequences explicitly.

The differences observed for the case of symmetric bidding hold more generally for the “win in  $l$ ” set and the associated “rule 3”: In “rule 3” of the transformed game, the lower item value realizes and captures that in “win in  $l$ ” the winning player has the higher signal and finds the lower item value realized. Crucially, there is – unlike in the auction game – no difference between the conditional and

---

<sup>5</sup>Due to this structure, deviations to “never win” and “win in  $\{l, h\}$ ” are not always possible.

unconditional value in the transformed game that players have to take into account. Let  $\gamma$  denote the set of both own actions  $b_i$  and beliefs about the opponents action  $b_j$  such that the auction is in the range “win in  $l$ ” or that equivalently “rule 3” is applied in the transformed game. Notably, the equilibrium strategy lies in this set.<sup>6</sup> Then, it holds that the conditional value of the item,  $E[W^*|win, \gamma, x_i] = x_i - 3$ , is lower than the unconditional one,  $E[W^*|\gamma, x_i] = x_i$ , in the auction game. This difference highlights that an adverse selection problem is present in the auction game. If bidders ignore this and bid  $b > -3$ , they lose money on average.<sup>7</sup> This distinction between conditional and unconditional expectation is, however, obsolete in the transformed game,  $E[W^*|win, \gamma] = E[W^*|\gamma] = W_l$ , since the lower conditional value of item is already incorporated in “rule 3” independently of who wins the item.<sup>8</sup>

Crucially, “rule 3” makes the conditional item value explicit, but at the same time the special auction rules of the transformed game fully reflect the underlying structure of the auction game. For this reason, best response functions in the two games describe the same optimal behavior when abstracting from boundary signals and assuming that players – as is true in equilibrium – bid type-independently in the auction game. Hence, by construction, equilibrium bids in the transformed game coincide with values of the equilibrium bid function in the auction game. Additionally, players in the two games face identical uncertainty, so that more general risk preferences than risk neutrality do not change the equivalence of the equilibrium strategies. Further important consequences are that noise as modeled for example in Quantal Response Equilibrium cannot account for differences

---

<sup>6</sup>Similarly,  $\alpha$  ( $\beta$ ) would denote that action and belief are such that the auction is in the range “win in  $\{l, h\}$ ” (“never win”) or that “rule 1” (“rule 2”) is applied in the transformed game.

<sup>7</sup>Some researchers define the WC as deviations from equilibrium bidding with less than normal profits (Crawford and Iriberri, 2007, CL). Kagel and Levin (1986) associate the WC with bids in excess of the conditional value since this entails negative profits *on average*, a more stringent definition. Empirically, we can analyze the differences between treatments in many ways, we report them both in terms of actual losses as well as differences from equilibrium. Kagel and Levin (1986) use the difference in strategic discounting between common-value and private-value settings as an indicator of the adverse selection problem. They show that for two players there is no difference in strategic discounting. This remains true in our setting as  $-8$  remains the equilibrium strategy when signals determine the private value of bidders. For our purposes, however, this equilibrium comparison obstructs the view on important out-of-equilibrium effects of adverse selection, as still reflected by the difference of the conditional and the unconditional item value.

<sup>8</sup>Notably, as pointed out by a reviewer, there is no such distinction for the “win in  $\{l, h\}$ ” set and “rule 1”. For this set, the conditional and unconditional value are the same both in the auction game,  $E[W^*|win, \alpha, x_i] = E[W^*|\alpha, x_i] = x_i$ , and the transformed game,  $E[W^*|win, \alpha] = E[W^*|\alpha] = (W_l + W_h)/2$ . This is true since the chance of the winning player that either the low or the high item value realizes is equally 50% (as when not winning at all) in both games. For “never win” or “rule 2” a conditional expectation on winning cannot be calculated.

between the two games, neither can the fact that the equilibrium strategy is at the lower end of the action space,  $b = -8$  (see also Charness and Levin 2009).<sup>9</sup>

**Proposition 1.** *The unique Nash equilibrium in the transformed game for both players is to bid  $b^e = -8$ .*

*The Nash equilibrium relative bid function for both players in the auction game is denoted  $b^e = b^*$ . For signals  $x \in [46, 228]$ , any  $b^*$  takes the value*

$$b^*(x) = -8. \quad (1)$$

**Proof.** See appendix A.3. ■

Note that the transformation generates common knowledge of the possible values of the item while the auction game’s signal structure prevents this. Therefore, signals close to the end-point value of 25 reveal the value of the item fully to one player, providing incentives not to bid according to equation 1. For signal values up to 46, subjects’ optimal bid function can be influenced through higher-order beliefs by those incentives as detailed in appendix A.3. We disregard the few observations that fall in this small range when analyzing our data and further discuss the matter in section 3.5. Outside of this range, the differences in the information structure between the two games do not influence equilibrium behavior. Notably, the revelation of the item value through signals close to the upper bound of 225 creates further incentives for bidding low and thus does not change the equilibrium strategy.

Overall, the comparison of the auction and the transformed game allows us to identify how the complexity of *conditional reasoning* affects the WC. We do not conceptualize conditional reasoning in a formal model since our focus is not on the cognitive process of the subject but on the objective challenges posed by the game. Nonetheless, it is important to point out that *conditional reasoning* refers – in our understanding – to thinking and conditioning on hypothetical situations that are induced by the *game structure* and that possibly put the focus on particular states of the world such as the event of winning the auction. Under this definition, we do not subsume the formation of beliefs like it occurs even in the transformed game when subjects form beliefs about the opponents’ behavior and “condition” on those beliefs to best respond. Although the expected value in the transformed

---

<sup>9</sup>Charness and Levin (2009) implement a corner equilibrium as well as an internal equilibrium and find no qualitative difference between them. In our auction game, no such simple modification is available because allowing for lower bids gives rise to a mixed strategy equilibrium. Importantly, we are mainly interested in treatment differences between similarly structured games, not absolute bid levels in individual games.

game still depends on beliefs, no further information about a state of the world can be inferred under this standard belief formation. In our definition, we follow the literature which emphasizes the problem of conditional reasoning exclusively in common value auctions, jury voting, persuasion, etc. (Charness and Levin 2009; Esponda and Vespa 2014).

## 2.2 Experimental Treatments and Procedures

The games implemented in the experiment differ along two dimensions. The first relates to conditional reasoning, which subjects have to deal with in the auction game but not in the transformed game. The second dimension addresses belief formation and varies the extent to which subjects have to form beliefs about their opponent. The difficulty of belief formation has been the subject of the large level- $k$  literature which provides models and plenty of evidence of heterogeneous and inconsistent beliefs (Crawford et al., 2013). We provide two different manipulations: While the “*full* belief manipulation” represents the more conventional approach to manipulate beliefs, the “*partial* belief manipulation” is more explorative and tries to shed some light on why forming beliefs is problematic.

In the *full* belief manipulation, subjects are confronted with naïve computer opponents, whose strategies are known. In sharp contrast to facing fellow human opponents, the need to form beliefs and to cope with strategic uncertainty is fully removed. The subjects are informed that the computer follows the naïve strategy  $b^C = 0$ . In absolute terms, this implies that it bids according to the signal or the mean value of the item, respectively. In the experiment, subjects have to round their bids to one cent of a unit. The best response is thus  $BR(b^C) = -5.99$  (“win in  $l$ ”). We deliberately do not implement a more complex or realistic strategy for the computer opponents due to their role in the next manipulation. Furthermore, subjects do not necessarily have to be able to best respond to complex belief distributions against human opponents either. When subjects realize in a first step that underbidding by  $BR(b^C) = -5.99$  is the best response to naïve play, they might recognize the equilibrium strategy – which in turn is the best response to  $-5.99$  – in a second step.

In the *partial* belief manipulation, subjects face human opponents subsequently to interacting with the computer without receiving any payoff feedback in-between. The preceding encounter with a deterministically acting, naïve opponent already places subjects in a simple scenario providing a first basic belief how opponents may act. Subjects can extend this simple scenario to craft a belief about human

|                | Sequence of games      |                                       |                     |                     |                          |
|----------------|------------------------|---------------------------------------|---------------------|---------------------|--------------------------|
|                | Part I                 |                                       | Part II             |                     |                          |
| Treatment      | 1                      | 2                                     | 3                   | 4                   | <i>Game identifiers:</i> |
| $\mathcal{AH}$ | <b><math>AH</math></b> | $AC_{\mathcal{AH}}$                   | $TH_{\mathcal{AH}}$ | $TC_{\mathcal{AH}}$ | $A$ auction game         |
| $\mathcal{TH}$ | <b><math>TH</math></b> | $TC_{\mathcal{TH}}$                   | $AH_{\mathcal{TH}}$ | $AC_{\mathcal{TH}}$ | $T$ transformed game     |
| $\mathcal{AC}$ | <b><math>AC</math></b> | <b><math>AH_{\mathcal{AC}}</math></b> | $TC_{\mathcal{AC}}$ | $TH_{\mathcal{AC}}$ | $H$ human opponent       |
| $\mathcal{TC}$ | <b><math>TC</math></b> | <b><math>TH_{\mathcal{TC}}</math></b> | $AC_{\mathcal{TC}}$ | $AH_{\mathcal{TC}}$ | $C$ computer opponent    |

*Notes:* In order to distinguish games by the treatment they belong to, we add a treatment subscript, e.g.  $AH_{\mathcal{AC}}$ , whenever it is not the first game in the treatment. Our analysis focusses mainly on the six games in boldface.

Table 1: Sequence of games in the four treatments.

opponents featuring strategic uncertainty. Interestingly, this setting provides a valid alternative belief formation manipulation to the single-person decision problem of facing computer opponents (see ILN, Costa-Gomes and Shimoji, 2015). While our manipulation clearly provides subjects with a basic scenario, the idea that this scenario helps subjects to play against humans rests on the assertion that the computer scenario is simple enough for subjects to gain a better understanding of the game which in turn should improve bids against humans.<sup>10</sup>

Our manipulations of conditional reasoning and – full – belief formation provide us with four basic games: the *auction game* either played against *human*,  **$AH$** , or *computer* opponents,  **$AC$** , and the *transformed game* either played against *humans*,  **$TH$** , or *computer opponents*,  **$TC$** . We use a within- and between-subject design, in which all subjects play all four different basic games in different sequences. Table 1 illustrates our four treatments. The treatment name is derived from the first game in each treatment. Each treatment is divided in parts I and II. The  $\mathcal{A}$  treatments start with the auction games ( $A$ ) in part I and have the transformed games ( $T$ ) in part II. In the  $\mathcal{T}$  treatments, this sequence is reversed. Within each part of these  $\mathcal{H}$  treatments, the opponents switch from human ( $H$ ) to computer opponents ( $C$ ). Subjects are instructed before each particular game. Hence, they know about the computer opponent and its strategy only after they have finished the initial game. In the  $\mathcal{C}$  treatments, this switch is reversed from  $C$  to  $H$ . Hence, subjects face human after computer opponents ( $\star H_{\star C}$ ).<sup>11</sup> Thus, the games  **$AH_{\mathcal{AC}}$**  and  **$TH_{\mathcal{TC}}$**

<sup>10</sup>If this assertion is not true, one could conjecture that the preceding encounter against computerized opponents might even worsen subjects' play or at least not improve it: subjects could just imitate the computer, playing 0, or be confused because they perceive the computer behavior as unrepresentative and unhelpful for bids against humans. Although we cannot a priori rule out that subjects' play could be worsened by our manipulation, we strongly conjectured otherwise and empirical evidence supports this conjecture.

<sup>11</sup>An expression with  $\star$  refers to both  $A$  and  $T$  games, e.g.  $\star H$  stands for the two games  $AH$  and  $TH$ .  $H$  alone refers generally to human opponents (including e.g.  $AH_{\mathcal{AC}}$ ).

capture our partial belief manipulation.

Our design generates data for clear between-subject comparisons and also allows for rich within-subject analyses. The six games in boldface allow us to disentangle the effects of conditional reasoning and belief formation on game play using a between-subject analysis. The remaining games in part I,  $AC_{\mathcal{AH}}$  and  $TC_{\mathcal{TH}}$ , allow us to analyze within-subject bid transitions from human to computer opponents. Part II provides further within-subject data on the learning transfer between auction and transformed games and vice versa as analyzed in appendix B.2. Data from all 16 games will be used to quantify the effect of cognitive complexities in a regression analysis. Overall, our interest is not so much in absolute levels of bids but in relative treatment differences between games.

In all treatments, the general instructions and the instructions for the games are read out aloud. We do not provide a control questionnaire because meaningful questions might highlight the adverse selection problem underlying the games. Instead, *frequently asked questions* that summarize main points of the games are read aloud. These FAQs have been generated based on trial sessions. Subjects play each specific game for three consecutive periods against randomly rematched subjects or the computer. Subjects are informed that they will first make all 12 decisions in the experiment before receiving any feedback.<sup>12</sup> We deliberately rule out that subjects get any payoff information after each game to illuminate the mechanism behind the WC, undisturbed from learning through feedback. It has been shown that experiences with losses and simple learning strategies enable subjects to eventually avoid the WC. Crucially, they seem to do this without overcoming the underlying cognitive complexities insofar as they do not transfer knowledge to similar situations in the future (see Kagel and Levin, 2002, p. 337). Although it is interesting in itself to analyze how subjects adapt in the long run even without a deep understanding, our core interest is in (a) whether people’s failure to *overcome* certain cognitive complexities causes the WC in the first place and in (b) which complexity is to blame. Focussing on initial responses allows tackling these questions and moreover enables a meaningful within-subject analysis. At the end of our analysis, we will provide a robustness check analyzing the influence of the two complexities on the rate and degree of convergence when feedback is provided (see section 3.4).

The experiments were conducted at the University of Mannheim in Spring and

---

<sup>12</sup>Because subjects do not receive any feedback after playing one period, in principle, it would have been possible to just implement one period per game. However, implementing three periods allows us to see whether subjects consistently play the same strategy across three periods for different values of the signal.



Autumn 2014. Overall, 12 sessions with 10 to 22 subjects in each session were run. In total, 182 subjects participated.<sup>13</sup> Participants received a show-up fee of 4€. We used “Taler” as an experimental currency where each Taler was worth 0.50€. Subjects received an initial endowment of 8 Taler in each of the two parts from which losses were subtracted and to which gains were added. Participants that made losses in both parts still kept their show-up fee, following standard procedures as implemented by Kagel and Levin (1986) and ILN. Sessions lasted on average 60-75 minutes and subjects earned on average 14.40€. Instructions and FAQs are reprinted in appendices B.4 and B.5.

### 2.3 Formalizing the Impact of Cognitive Complexities

The primitive of our approach is the focus on the objective game characteristics rather than a model of subjects’ reasoning. We provide a brief formalization of the quantitative impact on behavior caused by the objective game complexities studied here. If two game situations  $G \in \{G_0, G_1\}$ , differ only in the need to engage in one particular cognitive complexity  $D$ , we propose to evaluate this complexity based on the difference it generates in behavior between these games. For example,  $G_0 = TH$  and  $G_1 = AH$  differ only in the need to engage in conditional reasoning. We describe behavior  $s \in \mathcal{S}$  as a probability measure over the action space of  $G$  or as a statistic thereof. We define the distance between observed behavior  $s$  and a benchmark behavior such as equilibrium  $s^e$  as  $\tilde{\mu}(s, s^e)$ , where  $\tilde{\mu}(\cdot, \cdot) : \mathcal{S} \times \mathcal{S} \mapsto \mathbb{R}$  is an appropriately chosen divergence or metric.<sup>14</sup>

We define  $D$ ’s impact on behavior as

$$\Delta\tilde{\mu}(D; G_0) \equiv \tilde{\mu}(s(G_1), s^e) - \tilde{\mu}(s(G_0), s^e), \quad (2)$$

and can interpret  $D$  as follows. If the presence of  $D$  keeps behavior further away from equilibrium play (as we hypothesize for conditional reasoning),  $\Delta\tilde{\mu}(D) > 0$ , we call it a *cognitive bad*. The presence of a *cognitive good* results in  $\Delta\tilde{\mu}(D) < 0$ . Such a good can be viewed as resulting from a simplification (or negative complexity) like, potentially, repeated game play or the substitution of probabilities with

---

<sup>13</sup>The experimental software was developed in z-Tree (Fischbacher, 2007). For recruitment, ORSEE was used (Greiner, 2004).

<sup>14</sup>While this use of the distance is convincing in our particular games, we understand that this principle is not universally valid, in particular, outside of the class of games with continuous strategy space. At the same time, the creation of a distance metric analogue to our concept on the basis of a ranking of expected payoffs or rationalizability would still be in line with our concept and could certainly be helpful. Also, in case the benchmark behavior or action spaces are different between  $G_0$  and  $G_1$ , an appropriate normalization can assure comparability.

frequencies in games with Bayesian reasoning (Gigerenzer and Hoffrage, 1995). For simplicity, we omit the dependence of  $D$ 's impact on circumstances ( $\cdot; G_0$ ) when it is of minor importance. When comparing TH and AH, for example, these circumstances include that subjects in  $G_0 = TH$  (as well as in  $G_1 = AH$ ) have to form beliefs.

Two activities  $D_1$  and  $D_2$  (for example, conditional reasoning and belief formation) are related to each other depending on the difference between the joint effect and the sum of the individual effects. Define

$$\Delta^2 \tilde{\mu}(D_1, D_2) \equiv \Delta \tilde{\mu}(\{D_1, D_2\}) - (\Delta \tilde{\mu}(D_1) + \Delta \tilde{\mu}(D_2)), \quad (3)$$

where  $\{D_1, D_2\}$  indicates the need to engage in both cognitive processes. Two activities exhibit *cognitive diseconomies* if their joint effect is larger than their summed individual effects,  $\Delta^2 \tilde{\mu}(D_1, D_2) > 0$ . They exhibit *cognitive economies* if this difference is negative. This definition adapts the idea of diseconomies of scope  $\Delta^2 C$  of multi-input cost functions,  $C(a_1, a_2) = C(a_1, 0) + C(0, a_2) + \Delta^2 C(a_1, a_2; 0)$ .

The strategy space in our games is a subset of the metric space of the real numbers. We summarize the action distribution in terms of relative bids with the mean statistic and can resort to the simple Euclidean distance as metric. We normalize to 1 the distance between uninformed random play  $\mu^u(s^u, s^e) = 1$  and equilibrium or optimal play  $\mu^e(s^e, s^e) = 0$ . In particular, we use a simple normalization

$$\mu(s, s^e) \equiv \left| \frac{\bar{b} - b^e}{\bar{b}^u - b^e} \right|, \quad (4)$$

where  $\bar{b}$  is the average behavior, and  $b^e$  and  $\bar{b}^u$  are equilibrium or optimal play and uninformed random play averages, respectively. In our games,  $b_{\star H}^e = -8$ ,  $b_{\star C}^e = -5.99$  and  $\bar{b}^u = 0$ .

As we will detail in the following, figure 6 illustrates how the game complexities in our design relate. For our study, we denote the activity of conditional reasoning as  $D_A$ . As the dotted lines in figure 6 indicate, it is the additional complexity in the auction game compared to the transformed game. We distinguish different types of cognitive complexities in the context of belief formation. First, we denote as  $D_B$  the forming of any first scenario or belief that is provided in  $\star H_{\star C}$  by the experience with the computer but not in  $\star H$  games. Independent of whether a basic scenario or belief is provided to subjects or not, any belief has to take strategic uncertainty into account,  $D_{SU}$ , whenever subjects face human opponents.

We can relate these complexities to our two belief manipulations. Our *full belief*

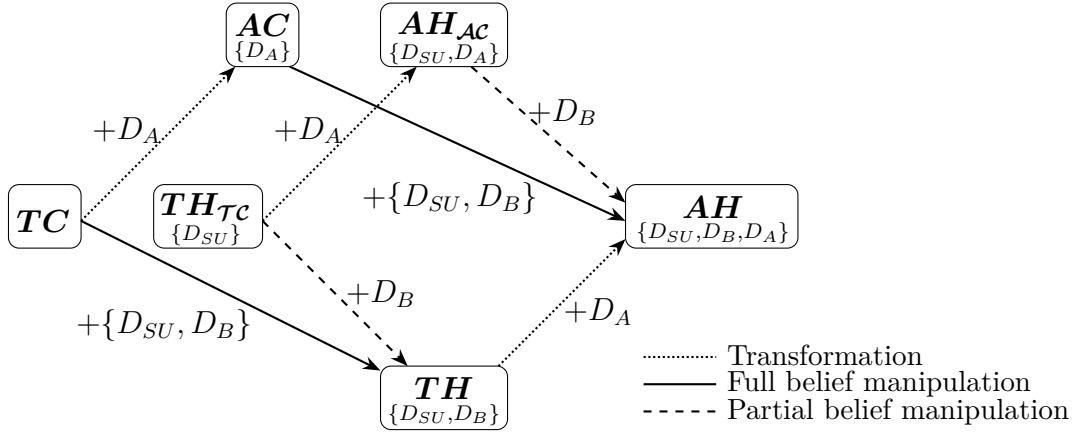


Figure 6: Differences between games with complexities added to  $TC$  in  $\{\cdot\}$ .

*manipulation* varies the necessity of forming beliefs against humans,  $D_H$ . This is done by comparing human settings  $\star H$  to computerized settings  $\star C$ , as indicated by the solid lines in figure 6. It entails adapting one's belief to an environment featuring strategic uncertainty  $D_{SU}$ , but as well entails forming a basic belief,  $D_B$ . Thus,  $D_H = \{D_B, D_{SU}\}$ . While it seems reasonable that strategic uncertainty  $D_{SU}$  complicates the situation compared to naïve computer play, it is a priori less clear whether subjects really have a problem of forming a basic first belief  $D_B$  in the case we find that  $D_H$  worsens play significantly. Inspired by the level- $k$  idea of iterated best responses, our underlying conjecture is that it might be easier in our games to start with a simple (e.g. naïve) belief and extent it to a more sophisticated one than building a belief from scratch. But, notably,  $D_B$  could be unproblematic because subjects easily craft a basic scenario themselves or because they can develop consistent beliefs featuring strategic uncertainty without the need of any first belief. Thus, the underlying idea of our *partial belief manipulation* is to provide some indication whether the forming of a first belief  $D_B$  actually provides an obstacle that can explain part of the overall effect of full belief formation,  $D_H$ . Since  $D_B$  cannot be an obstacle when playing against humans after computers in  $\star H_{\star C}$  but can be an obstacle when just playing against humans in  $\star H$ , the comparison of the two games enables us to identify  $D_B$ , as indicated by the dashed lines in figure 6.

Each individual cognitive complexity can be measured with at least two game comparisons since overall three manipulations are implemented. This reflects that the magnitude of an effect depends on circumstances,  $G_0$ . In the results section, we will analyze all depicted comparisons and also use a regression analysis to precisely quantify each cognitive complexity, controlling for interaction effects and thus

circumstances.

Our approach deliberately disregards the mechanisms behind subjects' inability to overcome a complexity  $D$ . We only assume that game complexity influences the distance of observed play to the equilibrium strategy. We remain agnostic as to whether it is the difficulty to form beliefs exactly as modeled in the level- $k$  model or the obstacle of conditional reasoning as specifically modeled in CE. It turns out that in our specific games neither of the two predicts differences between treatments. Appendix B.1 shows that except for the level- $k$  model with truthful level-0, these models predict the equilibrium bid of  $b^e = -8$ .

## 2.4 Hypotheses

The large literature on belief formation and conditional reasoning undoubtedly suggests that the two processes analyzed in this paper are cognitive biases. Additionally, they could interact and might exhibit cognitive diseconomies that distance behavior further from optimal play.

**Hypothesis 1 (Conditional reasoning):** *Due to the added need of conditional reasoning, subjects make higher bids and fall prey to the WC more frequently in the auction game compared to the transformed game, both with computer and human opponents as well as when facing human after computer opponents:  $\Delta\mu(D_A) > 0$ .*

**Hypothesis 2a (Full belief manipulation)** *Due to the added need of full belief formation, subjects make higher bids and fall prey to the WC more frequently in both games when playing against human opponents than when playing against computerized opponents:  $\Delta\mu(\{D_B, D_{SU}\}) > 0$ .*

**Hypothesis 2b (Partial belief manipulation):** *Due to the added need of forming a basic first belief, subjects make higher bids and fall prey to the WC more frequently with human opponents in both games if the game is played first compared to when it is played after the setting with computer opponents:  $\Delta\mu(D_B) > 0$ .*

**Hypothesis 3 (Magnitudes, cognitive diseconomies):** *The effects of  $D_H$  and  $D_A$  are different in magnitude:  $\Delta\mu(D_A) \neq \Delta\mu(D_H)$ . Similarly,  $\Delta\mu(D_A) \neq \Delta\mu(D_B)$ . Subjects make higher bids and fall prey to the WC more frequently when both  $D_A$  and  $D_H$  are present than predicted by  $\Delta\mu(D_A; TC)$  and  $\Delta\mu(D_H; TC)$  alone. In other words, we observe cognitive diseconomies  $\Delta^2\mu(D_A, D_H; TC) > 0$ . Weaker diseconomies are observed for  $D_B$  instead of  $D_H$ .*

### 3 Results

The following summary statistics and tests use the average bids and payoffs over the three periods of each specific game.<sup>15</sup> Only the percentage of winners incurring losses is calculated using the per-period information.

Since means and distributions only proxy for the plausibility of bids, we also report bids in four meaningful categories. They can account for the fact that equilibrium or optimal bids are different between the settings with human and computer opponents. These categories are determined by whether bids can be a valid best response. In the  $H$  games, the important thresholds are at  $b_i = -8, -5, -3$ . The first category is the equilibrium strategy, bidding  $b_i = -8$ .<sup>16</sup> The next threshold is the best response to a naïve strategy,  $b_j = 0$ , which we round up from the precise value  $b_i = -5.99$  to  $b_i = -5$  because some subjects only bid integer values. Finally, bidding  $b_i > -3$  is a weakly dominated strategy. Intuitively, whenever  $j$  bids very high values ( $b_j \geq 3$ ), no positive payoffs can be obtained, and any bid  $b_i \leq b_j - 6$  is a best response. Whenever positive payoffs can be achieved for lower bids of  $j$ , some strategy  $b_i < -3$  always leads to higher expected payoffs than bidding above  $-3$ . Overall, we think that bids  $b_i \in [-8, -5]$  represent plausible behavior. Bids  $b_i \in (-5, -3]$  might be a best response but only to some forms of fairly implausible beliefs; bids above are weakly dominated.

For the  $C$  games, a similar picture emerges in which we distinguish plausible behavior,  $b_i \in [-5.99, -5]$ , respectively, from implausible behavior with either  $b_i < -5.99$  or  $b_i > -5$ .<sup>17</sup>

#### 3.1 Conditional Reasoning (Hypothesis 1)

With respect to conditional reasoning, we are interested in comparisons between the auction and the transformed games in settings  $\star H$ ,  $\star C$ , and  $\star H_{\star C}$ . For all three comparisons, a higher percentage of winners in the auction game faces losses than

<sup>15</sup>In the few cases in which we drop observations due to a signal realization below 46, the average is over 2 bids.

<sup>16</sup>Given the empirical distribution of subjects' behavior, equilibrium play is not a best response, but it is close. In  $AH$ , bidding  $b_i = -7.97$  is the best response. In  $TH$ , bidding  $b_i = -7.99$  is the best response.

<sup>17</sup>When performing between-treatment comparisons below, we will use Fisher's exact tests that only rely on the two categories: plausible vs. implausible play. First, comparing human and computer settings with four categories is not desirable due to different category boundaries. Second, for similar within-subject comparisons, the McNemar's test is used that only relies on binary categories. Importantly, using four categories for testing whenever feasible leads to very similar results, suggesting that our analysis is robust at least with respect to the number of categories used.

Table 2: Summary statistics.

| Means<br>(Std. deviation)         |         | $A$                 | $T$                 | Wilcoxon<br>$p$ -value |
|-----------------------------------|---------|---------------------|---------------------|------------------------|
| $\star H$                         |         | $AH$                | $TH$                |                        |
|                                   | Bids    | −1.80<br>(2.63)     | −4.00<br>(2.61)     | 0.000                  |
|                                   | Payoffs | −0.56<br>(1.55)     | 0.55<br>(1.37)      | 0.001                  |
| $\star C$                         |         | $AC$                | $TC$                |                        |
|                                   | Bids    | −2.83<br>(3.65)     | −4.18<br>(2.90)     | 0.079                  |
|                                   | Payoffs | −0.12<br>(1.99)     | 0.37<br>(2.11)      | 0.099                  |
| $\star H_{\star C}$               |         | $AH_{\mathcal{AC}}$ | $TH_{\mathcal{TC}}$ |                        |
|                                   | Bids    | −2.62<br>(4.15)     | −4.64<br>(2.83)     | 0.024                  |
|                                   | Payoffs | −0.57<br>(2.46)     | 0.82<br>(1.56)      | 0.003                  |
| $\star H$ vs. $\star C$           | Bids    | 0.022               | 0.323               |                        |
| Wilcoxon<br>$p$ -value            | Payoffs | 0.033               | 0.434               |                        |
| $\star H$ vs. $\star H_{\star C}$ | Bids    | 0.166               | 0.173               |                        |
| Wilcoxon<br>$p$ -value            | Payoffs | 0.303               | 0.116               |                        |

*Notes:* The last column and the last rows report two-sided  $p$ -values of Wilcoxon rank sum tests that evaluate whether the distribution of bids and payoffs is different between games/treatments. Noteworthy, the difference in equilibria biases against observing a difference when comparing  $\star H$  (−8) vs.  $\star C$  (−5.99).

in the transformed game: 61% do so in  $AH$  but only 32% in  $TH$ ; 45% do so in  $AC$  but only 24% in  $TC$ . Finally, 52% do so in  $AH_{\mathcal{AC}}$  but only 21% in  $TH_{\mathcal{TC}}$ .<sup>18</sup> These outcomes follow from bidding behavior illustrated in table 2. Average bids in  $T$  are significantly lower and thus closer to the equilibrium or optimal behavior compared to  $A$ , in  $\star H$ ,  $\star C$ , and  $\star H_{\star C}$ . The differences in payoffs are also significantly different irrespective of the opponents. In  $A$ , subjects on average lose money while they gain in  $T$ . In all six games shown in table 2, bids and payoffs are highly negatively

<sup>18</sup>When discussing the extent of the WC, we refer to actual probabilities with which winners face losses. Table 3 and Figures A.1-A.4 nonetheless show that very similar percentages emerge when looking instead at those who bid above the conditional item value:  $b_i > -3$ . Especially for  $AH$ , the extent of the WC would even be higher under the latter measure (72% vs. 61%): bidding above  $-3$  may not lead to losses in case someone else bids even higher.

correlated with each other.<sup>19</sup>

Figure 7 reports subjects' bid distributions in the relevant games. The histograms in figure 7a and 7b reflect that subjects play lower bids more often in  $TH$  than they do in  $AH$ . Actually, the bidding behavior in  $AH$  gives the impression of normally distributed bids that do not reflect the equilibrium strategy of  $b_i = -8$  at all. In contrast, bidding behavior in  $TH$  at least partially reflects that the equilibrium is the lowest possible bid. For  $\star C$ , figures 7c and 7d show that a larger number of subjects is able to find the exact equilibrium when strategic uncertainty is absent. Finally, figures 7e and 7f show a similar difference between  $A$  and  $T$  when subjects play against humans after a preceding interaction with the computer,  $\star H_{\star C}$ .

Table 3 shows data on plausible behavior. It reveals that 39% of subjects (18 of 46) bid plausibly in  $TH$  while only 12% (6 of 50) do so in  $AH$  (Fisher's exact test,  $p = 0.004$ ).<sup>20</sup> Consistently, 57% (24 of 42) bid plausibly in  $TH_{TC}$  but only 39% (17 of 44) do so in  $AH_{AC}$  (Fisher's exact test,  $p = 0.083$ ). Moreover, a plausible action is played by 64% (27 of 46) in  $TC$  but only by 43% of subjects (19 of 44) in  $AC$  (Fisher's exact test,  $p = 0.090$ ).<sup>21</sup> Hence, even if subjects exactly know how their opponents act, conditional reasoning is still a problem, at least for some subjects.

Table 3: Category frequencies and percentages of bids.

|           | $[-8]$         | $(-8, -5]$      | $(-5, -3]$      | $(-3, 8]$ | Total  |
|-----------|----------------|-----------------|-----------------|-----------|--------|
| $AH$      | <b>0 (.00)</b> | <b>6 (.12)</b>  | 8 (.16)         | 36 (.72)  | 50 (1) |
| $TH$      | <b>6 (.13)</b> | <b>12 (.26)</b> | 12 (.26)        | 16 (.35)  | 46 (1) |
| $AH_{AC}$ | <b>1 (.02)</b> | <b>16 (.36)</b> | 6 (.14)         | 21 (.48)  | 44 (1) |
| $TH_{TC}$ | <b>4 (.10)</b> | <b>20 (.48)</b> | 9 (.21)         | 9 (.21)   | 42 (1) |
|           | $[-8, -6]$     | $[-5.99]$       | $(-5.99, -5]$   | $(-5, 8]$ | Total  |
| $AC$      | 1 (.02)        | <b>12 (.27)</b> | <b>7 (.16)</b>  | 24 (.55)  | 44 (1) |
| $TC$      | 1 (.02)        | <b>16 (.38)</b> | <b>11 (.26)</b> | 14 (.33)  | 42 (1) |

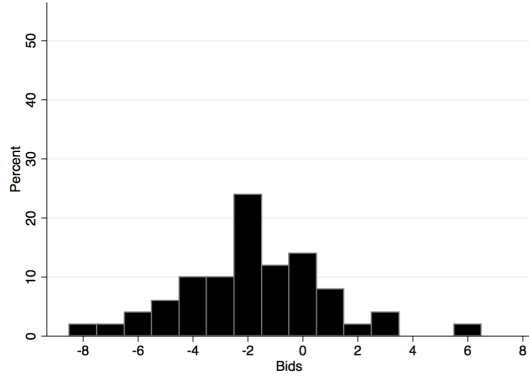
Notes: Plausible play in bold. Percentages in parenthesis. All in the text reported tests of differences in plausibility are  $2 \times 2$  tests of equality.

The analysis so far has focused on the six individual games illustrated in figure

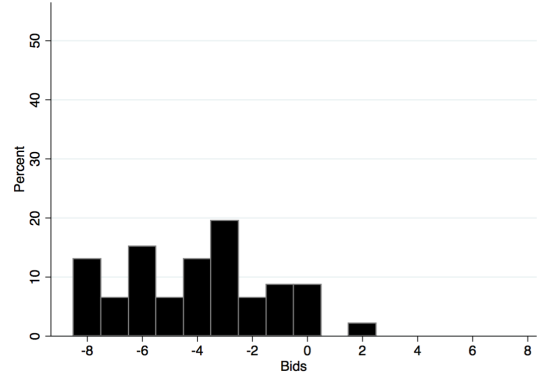
<sup>19</sup> $AH : -.77, TH : -.74, AC : -.77, TC : -.88, AH_{AC} : -.84, TH_{TC} : -.88$ , all  $p$ -values  $< 0.001$ .

<sup>20</sup>All reported tests in this paper are two-sided.

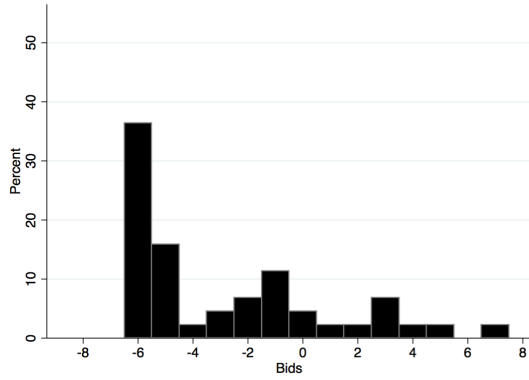
<sup>21</sup>In the computer setting, more subjects seem to understand the game and make lower but not excessively low bids under  $-5.99$ . This result speaks against the idea that potential differences in framing between the two games lead to more cautious, mechanically low play in the transformed game.



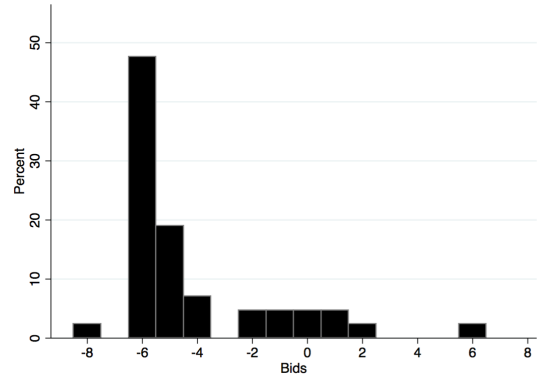
(a)  $AH$ ,  $N = 50$ .



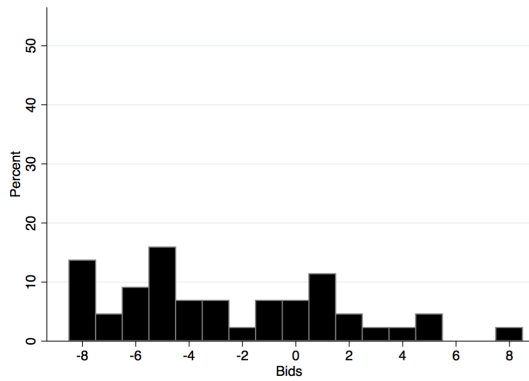
(b)  $TH$ ,  $N = 46$ .



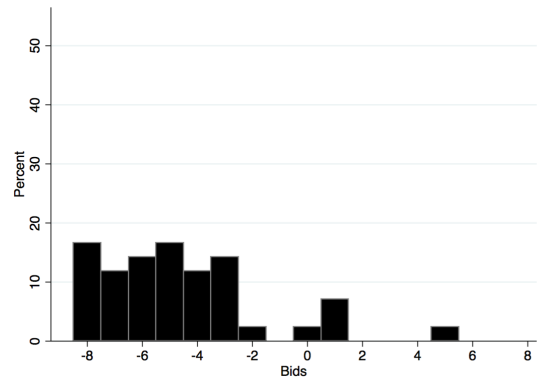
(c)  $AC$ ,  $N = 44$ .



(d)  $TC$ ,  $N = 42$ .



(e)  $AH_{AC}$ ,  $N = 44$ .



(f)  $TH_{TC}$ ,  $N = 42$ .

Figure 7: Histograms of bids.



6. Our design also allows us to study within-subject bid transitions between  $H$  and  $C$  as illustrated in table 4. Figures A.1 – A.4 graphically illustrate the individual data. Two important features are noteworthy. First, in  $\mathcal{AH}$  (but also in  $\mathcal{AC}$ ), 19 subjects play bids in the top-right cell, that is, higher than  $-3$  in  $H$  and higher than  $-5$  in  $C$ . This kind of bidding is less common in  $\mathcal{TH}$  and  $\mathcal{TC}$  where only 5 and 7 players do this. Second, in  $\mathcal{AH}$ , of those 27 subjects playing plausibly in  $C$ , 17 or 63% bid above  $-3$  in  $H$ , a weakly dominated strategy. Only 11 out of 37 (30%) do this in  $\mathcal{TH}$ . This suggests that beyond the ability to best respond, the need to reason conditionally increases the difficulty of belief formation. In  $\mathcal{AH}$  compared to  $\mathcal{TH}$ , relatively more subjects who are in  $C$  able to best respond fail to form adequate beliefs about human opponents and to best respond to them. In the  $\mathcal{C}$  treatments when the simpler  $\star C$  games are played first we do not observe such a difference.

Clearly, the results so far indicate that conditional reasoning is a cognitive bad as it distances behavior from optimal play,  $\Delta\mu(D_A) > 0$ . We relegate the discussion of the magnitudes and its comparisons across complexities to section 3.3 where we can base it on the regression results. Table A.1 on page 48 calculates the magnitudes for various effects and environments based on the average bids seen until now.

**Result 1:** For all three settings,  $\star H$ ,  $\star C$ , and  $\star H_{\star C}$ , we find that, with conditioning, subjects bid higher and avoid the WC less,  $\Delta\mu(D_A) > 0$ . Hence, even in CVA setting with human opponents, we find evidence that the difficulty of conditional reasoning is one reason behind the WC.

We find another consequence of the need to condition when we analyze subjects' behavior over the three periods of each game. When we test the equality of the distribution of bids in the first and the last period of each game, only 2 out of the 16 games of table 1 show significant differences. Subjects bid significantly closer to the equilibrium in the third compared to the first period only when the transformed game is played as the first game ( $\mathcal{TH}$ ,  $p=0.001$ , and  $\mathcal{TC}$ ,  $p=0.067$ ). Therefore, only without the need of conditional reasoning, subjects are able to improve their behavior even though they do not receive feedback.<sup>22</sup>

---

<sup>22</sup>Our central results regarding conditional reasoning and belief formation remain in general robust to considering first or third period bids instead of mean bids, although differences are less pronounced as single period data is naturally more noisy. Appendix B.2 provides further details.

Table 4: Frequency and percentage of bid transition by categories (Part I).

| $b_i \in [\dots]$        | $H$            |                 |                |                | Total           |
|--------------------------|----------------|-----------------|----------------|----------------|-----------------|
| $C$                      | $[-8]$         | $(-8, -5]$      | $(-5, -3]$     | $(-3, 8]$      |                 |
| $\mathcal{AH}$ treatment |                |                 |                |                |                 |
| $(-5, 8]$                | <b>0 (.00)</b> | <b>0 (.00)</b>  | 4 (.08)        | 19 (.38)       | 23 (.46)        |
| $(-5.99, -5]$            | <b>0 (.00)</b> | <b>1 (.02)</b>  | <b>0 (.00)</b> | <b>8 (.16)</b> | <b>9 (.18)</b>  |
| $[-5.99]$                | <b>0 (.00)</b> | <b>5 (.10)</b>  | <b>4 (.08)</b> | <b>9 (.18)</b> | <b>18 (.36)</b> |
| $[-8, -6]$               | <b>0 (.00)</b> | <b>0 (.00)</b>  | 0 (.00)        | 0 (.00)        | 0 (.00)         |
| Total                    | <b>0 (.00)</b> | <b>6 (.12)</b>  | 8 (.16)        | 36 (.72)       | 50 (1.0)        |
| $\mathcal{TH}$ treatment |                |                 |                |                |                 |
| $(-5, 8]$                | <b>0 (.00)</b> | <b>1 (.02)</b>  | 2 (.04)        | 5 (.11)        | 8 (.17)         |
| $(-5.99, -5]$            | <b>2 (.04)</b> | <b>2 (.04)</b>  | <b>5 (.11)</b> | <b>4 (.09)</b> | <b>13 (.28)</b> |
| $[-5.99]$                | <b>4 (.09)</b> | <b>9 (.20)</b>  | <b>4 (.09)</b> | <b>7 (.15)</b> | <b>24 (.52)</b> |
| $[-8, -6]$               | <b>0 (.00)</b> | <b>0 (.00)</b>  | 1 (.02)        | 0 (.00)        | 1 (.02)         |
| Total                    | <b>6 (.13)</b> | <b>12 (.26)</b> | 12 (.26)       | 16 (.35)       | 46 (1.0)        |
| $\mathcal{AC}$ treatment |                |                 |                |                |                 |
| $(-5, 8]$                | <b>0 (.00)</b> | <b>4 (.09)</b>  | 1 (.02)        | 19 (.43)       | 24 (.55)        |
| $(-5.99, -5]$            | <b>0 (.00)</b> | <b>3 (.07)</b>  | <b>4 (.09)</b> | <b>0 (.00)</b> | <b>7 (.16)</b>  |
| $[-5.99]$                | <b>0 (.00)</b> | <b>9 (.20)</b>  | <b>1 (.02)</b> | <b>2 (.02)</b> | <b>12 (.27)</b> |
| $[-8, -6]$               | <b>1 (.02)</b> | <b>0 (.00)</b>  | 0 (.00)        | 0 (.00)        | 1 (.02)         |
| Total                    | <b>1 (.02)</b> | <b>16 (.37)</b> | 6 (.14)        | 21 (.47)       | 44 (1.0)        |
| $\mathcal{TC}$ treatment |                |                 |                |                |                 |
| $(-5, 8]$                | <b>0 (.00)</b> | <b>3 (.07)</b>  | 4 (.10)        | 7 (.17)        | 14 (.33)        |
| $(-5.99, -5]$            | <b>1 (.02)</b> | <b>10 (.24)</b> | <b>0 (.00)</b> | <b>0 (.00)</b> | <b>11 (.26)</b> |
| $[-5.99]$                | <b>3 (.07)</b> | <b>6 (.14)</b>  | <b>5 (.12)</b> | <b>2 (.05)</b> | <b>16 (.38)</b> |
| $[-8, -6]$               | <b>0 (.00)</b> | <b>1 (.02)</b>  | 0 (.00)        | 0 (.00)        | 1 (.02)         |
| Total                    | <b>4 (.10)</b> | <b>20 (.48)</b> | 9 (.21)        | 9 (.21)        | 42 (1.0)        |

Notes: Plausible play in bold. Percentages in parenthesis.

### 3.2 Belief Formation (Hypothesis 2)

With respect to full belief formation, comparing  $\star H$  with  $\star C$ , we observe the following results: In  $AH$ , 61% of the auction winners face losses whereas only 43% do so in  $AC$ . In  $TH$ , 32% of the winning subjects face losses whereas in  $TC$  only 24% do so. Table 2 shows that in the  $A$  games subjects bid significantly lower and make significantly higher profits in  $\star C$  than in  $\star H$ . No statistical difference is observed for the  $T$  games. Note, however, that the difference in the optimal bid between  $C$  ( $-5.99$ ) and  $H$  ( $-8$ ) biases against observing a difference.

In contrast, measures of plausible play can be compared better, leading to a cleaner measure. Judging by the categories of table 3, 6 out of 50 subjects, or 12%, behave plausibly in  $AH$  while 19 out of 44, or 43%, do so in  $AC$  (Fisher’s exact test,  $p < 0.001$ ). In  $TH$ , 18 out of 46 subjects, or 39%, behave plausibly while 27 out of 42, or 64%, do so in  $AC$  (Fisher’s exact test,  $p = 0.011$ ).

**Result 2a:** We find that basic belief formation and strategic uncertainty jointly provide an additional obstacle for avoiding the WC both in the auction and in the transformed game,  $\Delta\mu(\{D_B, D_{SU}\}) > 0$ .<sup>23</sup>

The strong impact of the removal of both strategic uncertainty,  $D_{SU}$ , and basic belief formation,  $D_B$ , may not be surprising. However, within-subject analyses of the  $\mathcal{AH}$  and  $\mathcal{TH}$  treatments in table 4 (figures A.1 and A.3) provide first evidence that strategic uncertainty alone may not be able to explain the differences between settings. In both cases, out of those respective 27 and 37 subjects that approximately best respond in  $C$ , still a substantial fraction of 17 and 11 players, respectively, plays a weakly dominated strategy in  $H$ . This suggests that, although many subjects best respond in  $C$ , they seem not to form (and best respond to) *equilibrium* beliefs in  $H$ . We see different potential explanations for subjects’ behavior. On the one hand, subjects may not perceive computer behavior as an accurate representation of human behavior and consequently form very particular beliefs of others bidding higher than the signal ( $b_j \geq 3$ ) that could rationalize even weakly dominated strategies as a best response. Or subjects may develop particularly complex beliefs – a full belief distribution – that, unlike the simple belief in  $C$ , overstrain their best-response abilities, leading to dominated play.

---

<sup>23</sup>Crucially, the partial belief manipulation discussed below also shows that result 2a cannot exclusively be driven by the possibility that subjects have different preferences for computerized opponents compared to human opponents, when they, for example, experience a higher “joy of winning” with human opponents than with computer opponents. Notably, unlike the full manipulation, the partial manipulation compares two settings of exclusively human interaction.

On the other hand, subjects may already have problems to form any basic belief, impeding any best-response logic. The remaining two analyses in this section bring more evidence on this point and, in our opinion, favor the latter possibility that subjects have problems to form any a basic and, thus, any belief.

Under partial belief manipulation, 52% of subjects winning the auction in  $AH_{AC}$  face losses while 61% of subjects do so in  $AH$ . Moreover, 21% do so in  $TH_{TC}$  and 32% in  $TH$ . Table 2 shows that bids differ in the expected direction between  $AH$  and  $AH_{AC}$ , but this difference is not significant ( $p = 0.166$ ). Importantly, table 3, however, shows that plausible bids below  $-5$  are more frequent in  $AH_{AC}$  than in  $AH$  (Fisher’s exact test,  $p = 0.004$ ).

The bid transition analysis in table 4 (figure A.2) shows further that – just like in the  $\mathcal{AH}$  treatment – numerous subjects place bids in the top-right cell in  $\mathcal{AC}$  as well. More interestingly, in  $\mathcal{AC}$ , out of the 19 subjects approximately best responding in  $C$ , only 2 bid higher than  $-3$  in  $H$ . As mentioned, in  $\mathcal{AH}$ , out of 27 that play a best response in  $C$ , 17 bid higher than  $-3$  in  $H$ . Therefore, pairing the general ability to best respond with a little help in the belief formation – the computer providing a first basic belief – has a strong influence on the bids placed against human opponents.

The treatment  $\mathcal{TC}$  indicates that this effect is not due to the learning of conditioning during the preceding play in  $C$ . Table 2 shows that the absolute difference in bids between  $TH$  and  $TH_{TC}$  is 0.64 with a  $p$ -value of 0.173, very similar to the  $\mathcal{A}$  treatments.<sup>24</sup> The changes in the transition between  $C$  and  $H$  (table 4 and figure A.4) are similar to the  $\mathcal{A}$  treatments.<sup>25</sup>

**Result 2b:** Although the observed difference in mean bids is not significant, playing first against computer opponents leads to significantly more plausible play against human opponents, suggesting that subjects already have problems to form a basic belief irrespective of strategic uncertainty,  $\Delta\mu(D_B) > 0$ .<sup>26</sup>

---

<sup>24</sup>Due to the overall lower bidding in the  $\mathcal{T}$  treatments, there is no significant difference between categories as depicted in table 4 (Fisher’s exact test,  $p = 0.134$ ). If we, however, only consider those subjects who at least approximately best respond,  $b_i \in [-5.99, -5]$ , in  $C$ , more subjects play plausibly in  $TH_{TC}$  than in  $TH$  (Fisher’s exact test,  $p = 0.056$ ).

<sup>25</sup>Unsurprisingly, playing against humans first does not help playing against the computer in  $AC_{AH}$  or  $TC_{TH}$ . No significant differences between  $AC_{AH}$  and  $AC$  arise (bids,  $p = 0.524$ ; payoffs,  $p = 0.825$ ; plausible play,  $p = 0.310$ ). The same is true for  $TC_{TH}$  compared to  $TC$  (bids,  $p = 0.135$ ; payoffs,  $p = 0.301$ ; plausible play,  $p = 0.281$ ).

<sup>26</sup>A potential concern with our partial belief manipulation is that improvements are not due to the previous computer interactions but to the experience from 3 preceding periods played in  $\star H_{\star C}$  compared to none in  $\star H$ . Our data allows us to provide a robustness check by comparing

Although the usefulness of the computer interaction for subsequent behavior against humans makes the explanation with particular or complex beliefs less plausible, we want to provide additional evidence on what subjects think when making their bidding decision in the auction game. For that purpose, we implement the auction games  $AH$  and  $AH_{AC}$  with an intra-team communication design as in Burchardi and Penczynski (2014). In this design, teams of two players communicate about the bidding decision in a way that subjects have an incentive to share their reasoning about the game. Appendix A.4 discusses the details of the design and the results.

The communication data supports the conclusions previously drawn from our experimental design: Subjects have problems with belief formation (and conditional reasoning<sup>27</sup>): The data shows that only 15% of subjects deliberate – either in fairly concrete terms or vaguely – the other team’s decision. Less sophisticated deliberations including e.g. a tendency for cautionary bidding and a reference to the unconditional mean value do not require subjects to form belief and are overall more frequent. Looking at the few subjects that deliberate about others’ behavior does not reveal any evidence that subjects have beliefs of particularly high bids. They seem to form complex beliefs but these beliefs do not necessarily impede them from best responding. Although we cannot ultimately rule out that subjects may have beliefs that they choose not to discuss, we do observe that those teams mentioning beliefs bid on average much closer to equilibrium ( $\Delta$ : -2.93), suggesting that our communication analysis captures meaningful motives. In addition, subjects deliberate about others’ behavior much more frequently against the computer, as expected.

Overall, we see this as an additional indication that subjects may actually have problems to form even a basic belief. Although we acknowledge that it is difficult to determine what subjects actually think, two independent approaches – our partial manipulation and our communication analysis – favor this idea.

---

behavior in the third, last round of  $\star H$  with the first period of  $\star H_{\star C}$  (the fourth period of the  $\star C$  treatments) instead of using averages. While this robustness check supports our findings in the  $\mathcal{A}$  treatments, it indeed shows a smaller difference in the  $\mathcal{T}$  treatments. The reason is simple. Subjects already play fairly well in  $TH$ . In addition, their learning from period 1 to 2 does not leave much space for further improvements. Notably, in  $AH$  – which is our game of ultimate interest – subjects do not learn. Therefore, our robustness check shows that a preceding encounter with a computer opponent improves subjects’ behavior in the  $\mathcal{A}$  treatments, independent of the measure on which we base the analysis.

<sup>27</sup>Although 56% of subjects acknowledge in their messages the two possible states of the world, fewer subjects understand their implications: Only 34% state the implied possible information signals of the other team and only 24% mention that bid ranges differ. Those teams acknowledging the two states bid significantly lower ( $\Delta$ : -2.46).

### 3.3 Conditioning vs. Belief Formation (Hypothesis 3)

To conclude our main analysis, we investigate which of the two complexities, the necessity to condition or the necessity to form beliefs, has a stronger impact on subjects' behavior and how they relate to each other. Quantifying the treatment differences is necessarily a more challenging task than just establishing qualitative differences. Despite these concerns, we note, however, that the two different analyses presented below lead to broadly similar conclusions.

We start by looking at conditional reasoning and *full* belief formation. Figure 8 shows the empirical cumulative distribution functions (CDF) of the first games in all four treatments, capturing the manipulation of beliefs,  $D_H$ , and conditional reasoning,  $D_A$  (see the solid and dotted lines in figure 6). For an undistorted comparison, we adjust bids in  $C$  games such that the optimal bid is always reflected by  $\tilde{b} = -8$ , the equilibrium bid in  $H$  games.<sup>28</sup> The leftmost distribution closest to equilibrium shows behavior in  $TC$  without either complexity. Then, no significant differences emerge when we compare playing against human opponents which requires belief formation,  $TH$ , to playing in an auction setting which requires conditional reasoning,  $AC$  (plausible play: 39% vs. 43%,  $p = 0.831$ ; adjusted bids:  $-4.00$  vs.  $-3.97$ ; payoffs:  $0.55$  vs.  $-0.12$ ,  $p = 0.289$ ). Only a Kolmogorov-Smirnov test – testing for first-order stochastic dominance – suggests a difference in the underlying distributions, namely that higher values are bid in  $TH$  compared to  $AC$  (adjusted bids  $p = 0.016$ ). The distribution of  $AH$  reflects clearly the behavior furthest from equilibrium.<sup>29</sup>

In  $TC$ , without both complexities, 24% of winners face losses, 32% do so in  $TH$  with the belief formation problem, 39% with the conditioning problem in  $AC$ , and 61% when facing both in  $AH$ . The difference in fractions of winners facing losses between  $TH$  and  $AC$  is not significant according to a proportion test ( $p = 0.399$ ).

Figure 9 shows the CDFs of bids in the four  $\star H$  and  $\star H_{\star C}$  games used to analyze the *partial* belief formation manipulation (see the dashed and dotted lines in figure 6). This time, the least complex game  $TH_{TC}$  involves only strategic uncertainty,

<sup>28</sup>This adjustment makes use twice of the normalization proposed in equation 4. First, we normalize bids in  $C$  using  $b^e = -5.99$  and then invert the normalization for  $H$  games in order to map normalized bids into  $[-8, 8]$ . Our interest in the quantitative comparison of magnitudes of effects motivates this adjustment at this point. The qualitative results established in previous sections are even stronger with this adjustment.

<sup>29</sup>Noteworthy, not adjusting bids leads to very similar results because the significance level of non-parametric tests between  $TH$  and  $AC$  does not change (bids:  $-4.00$  vs.  $-2.83$ ,  $p = 0.336$ ). Moreover, a Kolmogorov-Smirnov test with unadjusted bids is unable to detect a difference, which is even more in line with our general conclusion that there is no difference between conditional reasoning and belief formation.

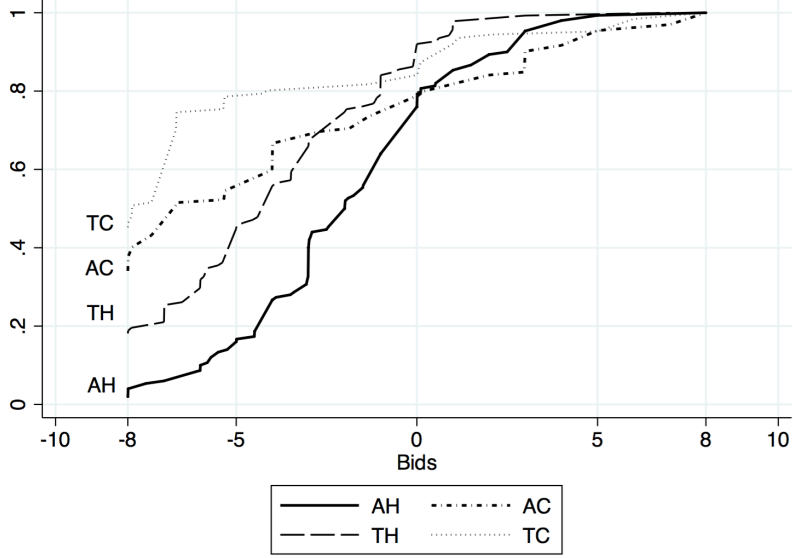


Figure 8: CDFs of subjects' adjusted bids: conditioning and full removal of belief formation

$D_{SU}$ , and results in the leftmost CDF closest to equilibrium. From there, moving to  $TH$  captures the difference of forming a first belief,  $D_B$ , and moving to  $AH_{AC}$  captures the difference of conditional reasoning. Here, the measurement of both added complexities already includes the interaction with strategic uncertainty. Interestingly, there is no significant difference between  $AH_{AC}$  and  $TH$  in plausible behavior or bids and only a marginally significant difference in payoffs (plausible play 39% vs. 34%,  $p = 1.000$ ; bids:  $-4.00$  vs.  $-2.62$ ,  $p = 0.148$ ; payoffs:  $0.55$  vs.  $-0.57$ ,  $p = 0.060$ ). Again, a Kolmogorov-Smirnov test suggests a difference in the underlying distributions ( $p = 0.062$ ). This time, however, bids in  $TH$  are lower than bids in  $AH_{AC}$ .<sup>30</sup>

In  $TH_{TC}$ , 21% of winners face losses, 32% do so in  $TH$ , 52% do so when the conditioning problem in  $AH_{AC}$  is activated, and 61% when facing both conditional reasoning and belief formation in  $AH$ . The difference between  $TH$  and  $AH_{AC}$  is significant according to a proportion test ( $p = 0.016$ ), highlighting that if we observe a difference in the partial manipulation at all, it is with respect to payoffs. Overall, considering both the partial and the full belief manipulation, the quantification of effects suggests that there is no clear difference in the effect magnitudes of conditional reasoning and belief formation. Most comparisons are insignificant.

<sup>30</sup>As noted earlier, having subjects play first in  $C$  seems to help those that are able to best respond, but it does not help those who have a more fundamental problem understanding the game. For this reason, the CDF of  $AH_{AC}$  is very similar to the CDF of the behavior in  $TH$  in the interval  $[-8, -5]$ . Further right, however, the former CDF approaches the CDF of  $AH$ .

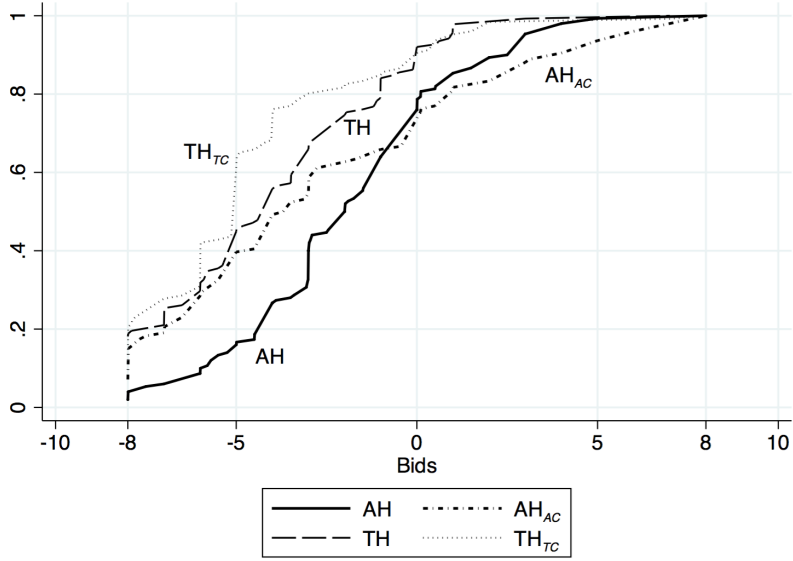


Figure 9: CDFs of subjects' bids: conditioning and partial removal of belief formation

While  $D_H$  is – if anything – a slightly higher obstacle than  $D_A$ ,  $D_B$  is – if anything – a slightly lower obstacle.

For a final overview, table 5 shows random-effects panel regressions in which the dependent variable is the bid. These regressions allow us to quantify the impact of the considered cognitive complexities  $\Delta\mu(\cdot)$  and at the same time to control for interaction effects and thus circumstances. All regressions include dummies that indicate the presence of the need to condition in auction games,  $D_A$ , the presence of the belief formation problem when playing against human opponents,  $D_H$ , and the interaction term of the two dummies,  $(\{D_A, D_H\})$ . Moreover, to additionally capture the partial belief formation, a dummy indicates whether subjects play against human opponents after playing against computer opponents,  $-D_B$ , removing  $D_B$ . Additionally, an interaction of the latter dummy with the conditioning problem is added,  $(\{D_A, -D_B\})$ .

While regressions 1, 3, and 5 only use part I data, regressions 2, 4, and 6 also include part II data. They additionally control for learning by including a part II-dummy, an interaction with the treatments  $\mathcal{TC}/\mathcal{TH}$  and a dummy for the last game in the treatment. Regressions 3 and 4 adjust bids as before for the fact that the optimal bid against the computer is different from the equilibrium bid against humans. This difference distortedly dampens the effect of the human opponents in 1 and 2. Finally, regressions 5 and 6 use normalized bids, allowing for an easy quantification of the effects in terms of  $\mu$  and  $\Delta\mu$ . Since the adjustment in 3 and



Table 5: Panel regression on bids

| Variables<br><i>Related complexity</i>      | Bids               |                    | Adjusted bids      |                    | Normalized bids   |                    |
|---|--------------------|--------------------|--------------------|--------------------|-------------------|--------------------|
|   | Part I<br>(1)      | Parts I&II<br>(2)  | Part I<br>(3)      | Parts I&II<br>(4)  | Part I<br>(5)     | Parts I&II<br>(6)  |
| Auction games                               | 1.49***<br>(0.46)  | 1.32***<br>(0.46)  | 1.89***<br>(0.61)  | 1.48***<br>(0.53)  | 0.24***<br>(0.08) | 0.18***<br>(0.07)  |
| $D_A$                                       |                    |                    |                    |                    |                   |                    |
| Human opponents                             | 0.86**<br>(0.37)   | 0.36<br>(0.29)     | 2.37***<br>(0.41)  | 1.57***<br>(0.33)  | 0.30***<br>(0.05) | 0.20***<br>(0.04)  |
| $D_H$                                       |                    |                    |                    |                    |                   |                    |
| Auction $\times$ Human                      | 0.63<br>(0.51)     | 1.02***<br>(0.36)  | 0.21<br>(0.59)     | 1.11***<br>(0.40)  | 0.03<br>(0.07)    | 0.14***<br>(0.05)  |
| $\{D_A, D_H\}$                              |                    |                    |                    |                    |                   |                    |
| Human after comp.                           | -1.18**<br>(0.51)  | -1.05**<br>(0.42)  | -1.33**<br>(0.53)  | -1.21***<br>(0.44) | -0.17**<br>(0.07) | -0.15***<br>(0.05) |
| $-D_B$                                      |                    |                    |                    |                    |                   |                    |
| Human a. comp. $\times$ Auction             | 0.01<br>(0.71)     | -0.14<br>(0.42)    | 0.04<br>(0.72)     | -0.13<br>(0.42)    | 0.00<br>(0.09)    | -0.02<br>(0.05)    |
| $\{D_A, -D_B\}$                             |                    |                    |                    |                    |                   |                    |
| Learning - Part II                          |                    | 0.88*<br>(0.47)    |                    | 1.11**<br>(0.54)   |                   | 0.14**<br>(0.07)   |
| Learning $\times \mathcal{TC}/\mathcal{TH}$ |                    | -2.56***<br>(0.79) |                    | -3.04***<br>(0.92) |                   | -0.38***<br>(0.11) |
| Lastgame                                    |                    | 0.56**<br>(0.24)   |                    | 0.68**<br>(0.29)   |                   | 0.08**<br>(0.04)   |
| Constant                                    | -4.61***<br>(0.29) | -4.39***<br>(0.27) | -6.04***<br>(0.38) | -5.67***<br>(0.34) | 0.24***<br>(0.05) | 0.29***<br>(0.04)  |
| N   | 364                | 728                | 364                | 728                | 364               | 728                |
| Subjects                                    | 182                | 182                | 182                | 182                | 182               | 182                |
| $R^2$ overall                               | .093               | .052               | .125               | .083               | .125              | .083               |

*Notes:* Panel random-effects regressions. The dependent variable is bids. For specifications (3)-(4), bids have been adjusted for settings with computer opponents to assure consistency of equilibrium bids (see footnote 28). For specifications (5)-(6), bids are normalized. Cluster-robust standard errors (subject level as level of independent observation) are provided in parentheses. \*\*\*, \*\* and \* indicate significance at the 1%, 5% and 10% level.

4 is already based on our normalization, coefficients in 5 and 6 can be gained by dividing those in 3 and 4 by 8 (with the only exception of the *constant*).

The results confirm that both complexities,  $D_A$  and  $D_H$ , individually lead to higher bids, significantly so when we adjust or normalize bids (regressions 3-6). Differences between the two coefficients are not significant in regressions 1 and 3-6 ( $Z$ -Test,  $p > 0.2$ ). Additionally, the interaction term always has a positive sign and is significant when we make use of all available data (regressions 2, 4, 6).<sup>31</sup> Moreover, providing a first basic belief by playing first against the computer significantly improves behavior.

Using all available data, effects can be easily quantified by looking at regression

<sup>31</sup>The results further support the idea that subjects bid lower in part II when they have played the transformed games in part I ( $\mathcal{TC}/\mathcal{TH}$ ) but not the auction games. Subjects' performance deteriorates in the very last game of the experiment, potentially due to exhaustion or loosened self-restraint. Part II results are in line with the more detailed analysis in appendix B.2.

6. The constant suggests that behavior without either complexity is relative close to the normalized equilibrium at 0,  $\mu(\bar{b}(TC), b^e) = 0.29$  (a bid of  $-5.67$  as regression 4 shows). As indicated by the dummy coefficients, both cognitive complexities deteriorate behavior by  $\Delta\mu(D_A; TC) = 0.18$  and  $\Delta\mu(D_H; TC) = 0.20$ , respectively. Moreover, their interaction intensifies this deviation with  $\Delta^2\mu(D_A, D_H; TC) = 0.14$ . This reflects the idea that full belief formation and conditioning exhibit cognitive diseconomies that reinforce the individual problems. All three effects lead to behavior in  $AH$  that is  $\Delta\mu(\{D_A, D_H\}; TC) = 0.52$  further away from benchmark play at  $\mu(\bar{b}(AH), b^e) = 0.81$  (a bid of  $-1.51$  in terms of regression 4).

Similar results emerge for the partial belief manipulation. Naturally, the process of finding a first belief is less complex than full belief formation, suggesting that a weaker effect should be observed. Reflecting our results so far, the problem of finding a first basic belief, irrespective of strategic uncertainty, is still substantial:  $\Delta\mu(D_B; TH_{TC}) = 0.15$ .<sup>32</sup> As expected, cognitive diseconomies between conditional reasoning and finding a first belief are smaller. They feature a non-significant value of  $\Delta\mu^2(D_A, D_B; TH_{TC}) = 0.02$ .

**Result 3:** Conditional reasoning is as high an obstacle to optimal play in an auction context as is belief formation,  $\Delta\mu(D_A) = \Delta\mu(D_H)$  and  $\Delta\mu(D_A) = \Delta\mu(D_B)$ . For partial belief formation we find insignificant diseconomies. Overall, however, there is significant evidence that the interaction of the two complexities intensifies the individual effects,  $\Delta^2\mu(D_A, D_H) > 0$ .

### 3.4 Repeated games with feedback

At the end of our analysis, we provide a robustness check and investigate how the presence of the two complexities under investigation influences the rate and the degree of convergence to equilibrium when subjects are provided with feedback. Unlike previous studies in which the WC was very persistent, we – ex ante – see the possibility of convergence to equilibrium in our games, because even our most complex game  $AH$  involves only two signals and two bidders and is, thus, simpler than usually studied CVAs. We therefore let subjects play the games repeatedly over 30 periods with immediate feedback between periods ( $AH^F$ ,  $TH^F$ ,  $AC^F$  and

---

<sup>32</sup>The effect of added conditional reasoning is  $\Delta\mu(D_A; TH_{TC}) = 0.30$ . Due to the starting point of  $TH_{TC}$ , this effect includes the added interaction with strategic uncertainty.  $\mu(\bar{b}(TH_{TC}), s^e) = 0.34$  can be calculated via *Constant + Human opponents + Human after comp.*. The transition to  $AH_{AC}$  can be calculated by *Auction games + Auction  $\times$  Human + Human a. comp.  $\times$  Auction*, resulting in  $\Delta\mu(D_A; TH_{TC}) = 0.30$ .

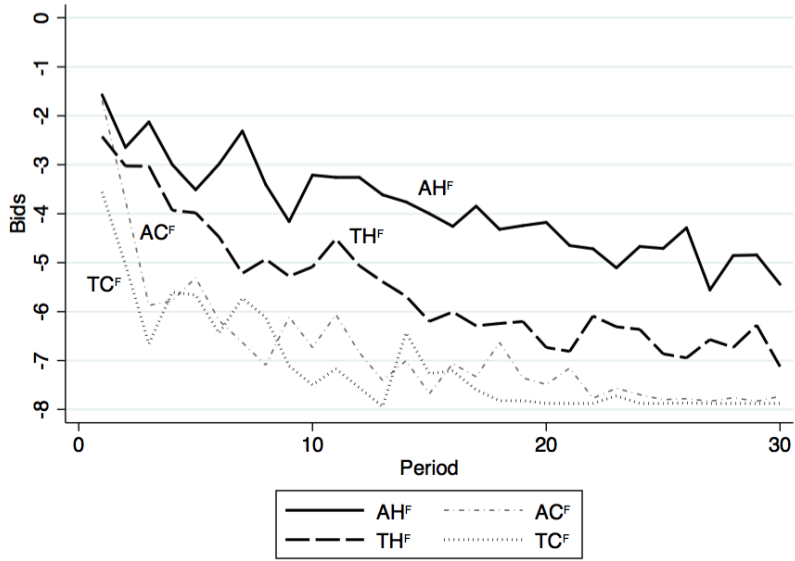


Figure 10: Mean bids in the repeated games with feedback over 30 periods.

$TC^F$ ).<sup>33</sup> After each period, subjects are randomly rematched in  $\star H^F$ . Our main question is whether conditional reasoning remains relevant when subjects learn to play the game over time. We, thus, conduct three sessions with 10 subjects each for  $AH^F$  and  $TH^F$ . Notably, we expect learning to be fundamentally different when subjects face a computer opponent that invariably uses the same strategy because the combination of feedback and the simplification of a computerized opponent creates a very simple learning environment.<sup>34</sup> For completeness, we, however, also implemented one session with 13 and 12 subjects, respectively, for the games  $AC^F$  and  $TC^F$ . For easy comparison, the  $\star C^F$  results will be given in adjusted bids.

With respect to the effect of conditional reasoning when playing against humans, figure 10 shows that the mean bids in  $AH^F$  are higher than in  $TH^F$  throughout all 30 periods. Non-parametric tests show a significant difference between the two games ( $p = 0.0495$ ,  $N = 3$ ). This also holds for periods 1-15 and 16-30 individually

<sup>33</sup>In line with Kagel and Levin (1986), we provide extensive feedback. Subjects are not only informed about the auction winner but also about the other person's signal and bid. Providing this feedback is an attempt to keep the informational setting constant between  $\star C^F$  and  $\star H^F$ . In  $\star C^F$ , subjects can often infer the computer's bid and signal even when only informed about who won the auction, while this would not be the case in  $\star H^F$ .

<sup>34</sup>This implies that differences between  $\star H^F$  and  $\star C^F$  could occur because of differences due to the underlying complexities or – in their absence – because of the complexities' different interaction with the learning from feedback. We leave it for future investigation to clarify this problem. Although our robustness check might, thus, not be suitable to replicate the fully-fledged analysis we provide for initial responses (e.g. quantification through our regression analysis), it can still illuminate how our complexities influence behavior over time and to what extent they matter even after 30 periods of play.

Table 6: Panel regression on bids over time ( $AH^F$  vs.  $TH^F$ )

| Variables                             | Periods            |                    |                    |                    |                    |
|---------------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|                                       | 1-30<br>(1)        | 1-30<br>(2)        | 1-15<br>(3)        | 16-30<br>(4)       | 1-30<br>(5)        |
| Auction games                         | 1.63***<br>(0.30)  | 1.24<br>(0.83)     | 0.69<br>(0.89)     | 2.16***<br>(0.23)  | 0.68<br>(0.89)     |
| Period                                | -0.12***<br>(0.03) | -0.13***<br>(0.05) | -0.22***<br>(0.06) | -0.05<br>(0.04)    |                    |
| Period $\times$ Auction Games         |                    | 0.02<br>(0.05)     | 0.10*<br>(0.06)    | -0.04<br>(0.05)    |                    |
| Period (1-15)                         |                    |                    |                    |                    | -0.22***<br>(0.06) |
| Period (1-15 $\times$ Auction Games)  |                    |                    |                    |                    | 0.10*<br>(0.06)    |
| Period (16-30)-Dummy                  |                    |                    |                    |                    | -3.35***<br>(0.89) |
| Period (16-30)-Dummy                  |                    |                    |                    |                    | 1.45<br>(0.89)     |
| Period (16-30)                        |                    |                    |                    |                    | -0.05<br>(0.04)    |
| Period (16-30) $\times$ Auction Games |                    |                    |                    |                    | -0.04<br>(0.05)    |
| Constant                              | -3.67***<br>(0.47) | -3.50***<br>(0.78) | -2.79***<br>(0.86) | -6.14***<br>(0.04) | -2.79***<br>(0.86) |
| N                                     | 1660               | 1660               | 825                | 835                | 1660               |
| Subjects                              | 60                 | 60                 | 60                 | 60                 | 60                 |
| $R^2$ overall                         | 0.159              | 0.161              | 0.010              | 0.106              | 0.168              |

*Notes:* Panel random-effects regressions using  $AH^F$  and  $TH^F$  data. The dependent variable is bids. For all specifications, bids have been adjusted for settings with computer opponents to assure consistency of equilibrium bids (see footnote 28). Cluster-robust standard errors (session level as level of independent observations) are provided in parentheses. A multilevel mixed-effects panel regression that accounts for variation within session by including random effects not only at the subject but also at the session level, leads to very similar results (Moffat 2016). \*\*\*, \*\* and \* indicate significance at the 1%, 5% and 10% level.

(both  $p = 0.0495$ ,  $N = 3$ ).<sup>35</sup>

The figure further shows that the initial difference is widened over time, suggesting that the learning proceeds quicker in  $TH^F$ . The panel regression – focusing on  $AH^F$  and  $TH^F$  data – in table 6 quantifies this perception. The basic specification

<sup>35</sup>Since subjects receive feedback, observations are only statically independent at the session level, leaving us with only three observations per treatment. Under these conditions, the above  $p$ -value of 0.0495 is minimal in the sense that all three  $AH^F$  observations are higher than the three  $TH^F$  observations. Similarly, both for  $AH^F$  as well as  $TH^F$ , average bids in periods 16-30 are lower than those in 1-15 in all three sessions (implying a minimal  $p$ -value of 0.1088). Since we run only one session for  $\star C^F$  treatments, comparing  $AH^F/TH^F$  to  $AC^F/TC^F$  with non-parametric tests is only possible when using individual observations. When doing so, differences are always significant ( $p < 0.01$ ), for 1-15, 16-30 and 1-30.

1 shows that the auction game features a significantly higher bid level and bids decrease in both games. Specification 2 adds an interaction term and, thus, controls for differences in slope. In this specification, no difference in the slope (and in the level) is observed. Specification 3, 4, and 5, however, distinguish the first (1-15) and the second part (16-30) of all periods. For periods 1-15, 3 and 5 show an initial difference between the two games that is, however, not statistically significant ( $\Delta_3 = 0.69$ ,  $\Delta_5 = 0.68$ ). Crucially, the learning path is significantly steeper for  $TH^F$ . This is consistent with subjects' learning – without feedback – from period one to three in  $TH$  but not in  $AH$ . The learning leads to considerable differences between the two games in later periods ( $\Delta_4 = 2.16$ ,  $\Delta_5 = 2.13 [= 0.68 + 1.45]$  in period 16). In later periods, however, learning proceeds – if anything – quicker in  $AH^F$ , as bidding in  $TH^F$  is already fairly close to the equilibrium bid of  $-8$ .

As expected, figure 10 reveals that the difference between a repeated computer interaction,  $\star C^F$ , and a repeated human interaction,  $\star H^F$ , leads to a fundamental difference in initial periods. While in periods 2 and 3 alone, subjects on average adjust their bids by  $-2.09$  ( $-1.55$ ) in  $AC^F$  ( $TC^F$ ), learning in  $\star H$  is much slower, with an adjustment of only  $-0.40$  ( $-0.31$ ) points in  $AH^F$  ( $TH^F$ ).<sup>36</sup> Notably, despite these fundamental differences between  $\star C^F$  and  $\star H^F$  in learning, initial differences without learning are – although less pronounced – roughly the same as before, with lowest bids for  $TC$  and highest bids for  $AH$  with  $AC$  and  $TH$  in between. The perspective of 29 additional periods may limit subjects' initial deliberations, potentially diminishing initial differences that are, in addition, based on fewer observations than in the main treatments.

Overall, our data clearly suggests that the need to engage in conditional reasoning is relevant for the learning process and still matters after 30 periods with full feedback. More generally, when looking at all four treatments, not only the rate of convergence but also the final degree of plausible and equilibrium play differs between treatments. In the last period, only 60% play plausible (26% play the equilibrium of  $-8$ ) in  $AH^F$ , while 90% (50%) do so in  $TH^F$ , 100% (69%) do so in  $AC^F$ , and 100% (75%) do so in  $TC^F$ .<sup>37</sup> Thus, our feedback analysis reveals that the observed difference without feedback translates to a meaningful difference in the rate and degree of convergence with feedback. Due to our simple environment,

<sup>36</sup>Even though only one session per  $\star C$  treatment was implemented, these differences are either significant or very close to significance with  $p$ -values between 0.043 and 0.138. Here,  $p$ -values are gained from a panel regression similar to specification 5 of table 6.

<sup>37</sup>Fisher exact tests reveal that the observed differences between  $AH^F$  and the other treatments are mostly significant:  $AH^F$  vs.  $TH^F$  – plausible:  $p = 0.015$ , equilibrium:  $p = 0.110$ ;  $AH^F$  vs.  $AC^F$  – plausible:  $p = 0.008$ , equilibrium:  $p = 0.016$ ;  $AH^F$  vs.  $TC^F$  – plausible:  $p = 0.009$ , equilibrium:  $p = 0.006$ .

we potentially observe less persistence in the WC than in other studies. But the deviation in average behavior from equilibrium behavior in  $AH^F$  – when both complexities are present – remains non-negligible.

### 3.5 Discussion

Most of our analysis is possible thanks to the transformation of the auction game that removes the need to condition. Certainly, this transformation also removes private information from the auction and establishes common knowledge of the possible values. Subjects thus never form higher-order beliefs about players with different information. While from a theoretical perspective this is clearly not a conditioning manipulation that leaves everything else the same, there are various empirical arguments that convince us that the only behaviorally relevant change is in the need to condition.

First, in the computer treatments, the games only differ in the need to condition because subjects do not need to form higher-order beliefs in the auction game. These treatments, however, yield similar differences as in the human opponent setting. Relatedly, our partial belief manipulation and our communication analysis provide independent evidence that many subjects have problems to form even a basic first belief. These subjects surely do not differentiate between opponents with different signals.

Second, notably, the effect of conditional reasoning  $\Delta\mu(D_A)$  can be quantified based on the comparison between the surely unproblematic computer treatments. Moreover, the belief formation effect  $\Delta\mu(D_H)$  can be quantified by comparing human and computerized opponents in the transformed game. Thus, only the interaction effect  $\Delta^2\mu(D_A, D_B)$  inevitably involves the auction game with human opponents. Thus, the diseconomies between conditional reasoning and belief formation might be overestimated if the mere possibility of higher-order beliefs in the auction game makes forming beliefs – even those that do not involve higher-order beliefs – in general more difficult. Evidence from the partial belief manipulation does not support this worry. Providing a simple first belief – without higher-order considerations – leads to a similar improvement in the auction as in the transformed game. If possible higher-order beliefs made belief formation more difficult in general, one would expect the partial belief formation to have a smaller effect in the auction game.

Third, in the range of signals considered in our data analysis, the equilibrium consists of strategies of constant relative bidding. In our experiments, subjects

indeed bid in accordance with this feature irrespective of the absolute signal realization.<sup>38</sup> In equilibrium and empirically, there is thus no reason to differentiate beliefs with respect to the signal value.

Fourth, despite our efforts to frame the games similarly, one might be concerned that differences in the games' framing could explain the observed behavioral differences. Importantly, any framing explanation must be valid for different types of players. After all, we do not only observe differences in play of weakly dominated strategies but also observe a sharp difference in plausible play and even equilibrium play. Assuming e.g. that confused subjects are more likely to anchor their decision differently between the two games (signal vs. lower item value) does not explain why many of those players become sophisticated and play the equilibrium strategy in the transformed game.

Fifth, while it is in principle possible to implement the transformed game as a fully strategically equivalent game, this is very complicated for subjects since instructions would involve rules in terms of both absolute and relative bids.<sup>39</sup> Then, subjects might not approach such a fully equivalent game in terms of the described rules but in terms of the intuitive standard rules of an auction. An advantage of our transformation is the generation of a distinct setting that is not perceived as a standard first price-auction, also because it lacks the standard signal structure.

Finally, one could manipulate instructions such that they explicitly explain the conditioning problem without manipulating the actual auction game. It is, however, difficult to provide instructions that explain conditioning without influencing sophistication in terms of belief formation. Overall, therefore, we believe that the transformation is very useful and its only behaviorally relevant variation lies in the need to condition.

As a starting point, our transformation requires an auction game that is simplified compared to more standard CVAs. We restrict the number of subjects to two and let binary signals be drawn without replacement. While we expect that

---

<sup>38</sup>When regressing the relative bids on absolute signals or on the mean value of the item for each of the four games, we only observe a significant positive effect for the transformed game with computer opponents. For the same game with human opponents and more importantly for both auction game settings, positive and negative coefficients arise and are always insignificant.

<sup>39</sup>Such a game would use the standard signal structure but replace the auction rule with rules set in terms of relative bids like in the transformed game. It follows that the description of the game setting would be in absolute bids, whereas the rules would use a relative perspective. Finally, profit calculations would again have to rely on absolute bids. Implementing and describing these changes of perspective would be very cumbersome and also constitute a major difference to the intuitive standard auction. A change of behavior would not be unambiguously attributable to the need of conditioning or to this change. Introducing common knowledge about the commodity's values allows us, however, to avoid a change between relative and absolute bids in the instructions.

belief formation and conditional reasoning continue to matter similarly in more typical auctions as both activities likewise remain indispensable to reach a best response in more complex auctions, it is an open question what exactly the impact and interaction of added complexities such as a larger number of players or a richer state space would be. To which extent is a richer state space a cognitive bad? Does it exhibit cognitive diseconomies with conditional reasoning? If answers to these questions were known, combining such elements in an extended framework of cognitive complexities could inform expectations for extensions of auction games as well as predictions in a large number of other games.

Focusing on conditional reasoning, our transformation can potentially be applied to other games and environments. In the strategic voting literature, players are conditioning on being pivotal in a jury decision (Feddersen and Pesendorfer, 1998; Guarnaschelli, McKelvey, and Palfrey, 2000). Using a computer experiment, Esponda and Vespa (2014) find that the cognitive difficulty of this operation might stand in the way of strategic voting.

After all, one might be tempted to turn to the behavioral models with our data. Recall that both the CE model as well as the level- $k$  model with uniform level-0 distribution applied to our setting predict the Nash equilibrium behavior  $b = -8$  (see appendix B.1). A level- $k$  model with truthful level-0 play rationalizes behavior by attributing 72% of players to level-0 types in the *AH* game. Only the transformed games exhibit the standard hump-shaped level- $k$  distribution and *TC* features a reasonable level average above 1.02. For this particular model, the observed difficulty of conditioning as well as the interaction between conditioning and belief formation might imply that higher levels are generally more difficult to reach and lower averages are to be expected. These potential dependencies highlight the value of considering complexities of a game and investigating their impact on behavior. Thus, our analysis can meaningfully complement studies that investigate particular models of reasoning such as Crawford and Iriberri (2007).

## 4 Conclusion

This study jointly analyzes two cognitive complexities associated with the winner’s curse: conditional reasoning and belief formation. First, we transform a common-value first-price auction in a way that subjects do not need to condition on hypothetical future events. Second, we remove the need to form beliefs by letting subjects play either against naïve computer opponents or against human opponents



subsequent to play against the naïve computer.

We provide a simple formalization of the impact of cognitive complexities on game play and can state the results as follows. Both activities, conditional reasoning and belief formation, constitute *cognitive bads* that significantly impair subjects' game play to a similar extent. Although their individual impact is limited, in combination they lead to widespread implausible behavior, exhibiting *cognitive diseconomies* as they interact and reinforce each other. Hence, adding the obstacle of conditional reasoning to the problem of belief formation – as it is done in many auctions – results in an extreme case of game-dependent sophistication. Our detailed experimental analysis allows us to reconcile the critique of Ivanov et al. (2010) of belief-based models with the numerous results in favor of them from other domains.

In our view, the focus on the elementary cognitive complexities is a useful complementary approach to the more common focus on behavioral models. Rather than explaining behavior, a better understanding of games' objective cognitive complexities and their effects could particularly improve the prediction across games. Investigating a wide range of games from this perspective – even in a meta-analysis of existing data – might be an insightful exercise for future research.

## Acknowledgments

We would like to thank Madhav Aney, Ayala Arad, Gary Charness, Vincent Crawford, Dirk Engelmann, Ori Heffetz, Nikos Nikiforakis, Ted O'Donoghue, Ariel Rubinstein, Larry Samuelson, Thomas Tröger and Emanuel Vespa as well as seminar participants in NYU Abu Dhabi, Cornell, Frankfurt, Heidelberg, Karlsruhe, Mannheim, Nuremberg, UC Santa Barbara, SMU Singapore, Tel Aviv, SABE Meeting 2014, ESA European Meeting 2014 (Prague), ESA North-American Meeting 2014 (Fort Lauderdale), theem 2015 (Konstanz), M-BEES 2015 (Maastricht), GAMES 2016 (Maastricht) for helpful comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- Avery, C. and J. H. Kagel (1997). Second-price auctions with asymmetric payoffs: An experimental investigation. *Journal of Economics & Management Strategy* 6(3), 573–603.
- Bazerman, M. H. and W. F. Samuelson (1983). I won the auction but don’t want the prize. *Journal of Conflict Resolution* 27, 618–634.
- Brandts, J. and G. Charness (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics* 14(3), 375–398.
- Brocas, I., J. D. Carrillo, S. W. Wang, and C. F. Camerer (2014). Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *The Review of Economic Studies* 81(3), 944–970.
- Burchardi, K. B. and S. P. Penczynski (2014). Out of your mind: Eliciting individual reasoning in one shot games. *Games and Economic Behavior* 84(0), 39–57.
- Camerer, C., S. Nunnari, and T. R. Palfrey (2015). Quantal response and nonequilibrium beliefs explain overbidding in maximum-value auctions. Working paper.
- Capen, E. C., R. V. Clapp, and W. M. Campbell (1971). Competitive bidding in high-risk situations. *Journal of petroleum technology* 23(06), 641–653.
- Charness, G. and D. Levin (2009). The origin of the winner’s curse: A laboratory study. *American Economic Journal, Microeconomics* 1(1), 207–236.
- Charness, G., D. Levin, and D. Schmeidler (2014). A generalized winner’s curse: An experimental investigation of complexity and adverse selection.
- Chou, E., M. McConnell, R. Nagel, and C. R. Plott (2009). The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two person “guessing” game. *Experimental Economics* 12(2), 159–179.
- Costa-Gomes, M. A. and M. Shimoji (2015). A comment on “Can relaxation of beliefs rationalize the winner’s curse?: An experimental study”. *Econometrica* 83(1), 375–383.

- Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature* 51(1), 5–62.
- Crawford, V. P. and N. Iriberri (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *Econometrica* 75(6), 1721–1770.
- Esponda, I. (2008). Behavioral equilibrium in economies with adverse selection. *The American Economic Review* 98(4), 1269–1291.
- Esponda, I. and E. Vespa (2014). Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics* 6(4), 180–202.
- Eyster, E. and M. Rabin (2005). Cursed equilibrium. *Econometrica* 73(5), 1623–1672.
- Feddersen, T. and W. Pesendorfer (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review*, 23–35.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Gigerenzer, G. and U. Hoffrage (1995). How to improve bayesian reasoning without instruction: frequency formats. *Psychological review* 102(4), 684.
- Glazer, J. and A. Rubinstein (2004). On optimal rules of persuasion. *Econometrica* 72(6), 1715–1736.
- Goeree, J. K. and T. Offerman (2002). Efficiency in auctions and public information in common values: An experimental study. *American Economic Review* 92(3), 347–361.
- Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*. 79–93.
- Grosskopf, B., Y. Bereby-Meyer, and M. Bazerman (2007). On the robustness of the winner’s curse phenomenon. *Theory and Decision* 63, 389–418.
- Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000). An experimental study of jury decision rules. *American Political Science Review*, 407–423.

- Ivanov, A., D. Levin, and M. Niederle (2010). Can relaxation of beliefs rationalize the winner’s curse?: An experimental study. *Econometrica* 78(4), 1435–1452.
- Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory* 123(2), 81 – 104.
- Jehiel, P. and F. Koessler (2008). Revisiting games of incomplete information with analogy-based expectations. *Games and Economic Behavior* 62(2), 533–557.
- Kagel, J. H. and D. Levin (1986). The winner’s curse and public information in common value auctions. *American Economic Review* 76(5), 894–920.
- Kagel, J. H. and D. Levin (2002). Bidding in common-value auctions: A survey of experimental research. *Common Value Auctions and the Winner’s Curse* 1, 1–84.
- Levin, D., J. Peck, and A. Ivanov (2016). Separating bayesian updating from non-probabilistic reasoning: An experimental investigation. *American Economic Journal: Microeconomics* 8(2), 39–60.
- Levin, D. and J. P. Reiss (2012). Could we overcome the winner’s curse by (behavioral) auction design? *Mimeo*.
- Lind, B. and C. Plott (1991). The winner’s curse: Experiments with buyers and sellers. *American Economic Review* 81(1), 335–346.
- Louis, P. (2015). The barrel of apples game: Contingent thinking, inference from others’ actions, and strategic heterogeneity.
- McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior* 10(1), 6–38.
- Moffat, P. G. (2016). *Experiments - Econometrics for Experimental Economics*. Palgrave.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review* 85(5), 1313–1326.
- Nagel, R. and F. F. Tang (1998). Experimental results on the centipede game in normal form: an investigation on learning. *Journal of Mathematical Psychology* 42(2), 356–384.

- Ngangoue, K. and G. Weizsäcker (2015). Learning from unrealized versus realized prices.
- Osborne, M. J. and A. Rubinstein (1994). *A Course in Game Theory*. MIT Press.
- Roll, R. (1986). The hubris hypothesis of corporate takeovers. *Journal of business*, 197–216.
- Stahl, D. O. and P. W. Wilson (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10(1), 218–254.
- Tversky, A. and D. Kahneman (1986). Rational choice and the framing of decisions. *The Journal of Business* 59(4), pp. S251–S278.

# A Appendix

## A.1 Figures: Bid Transitions

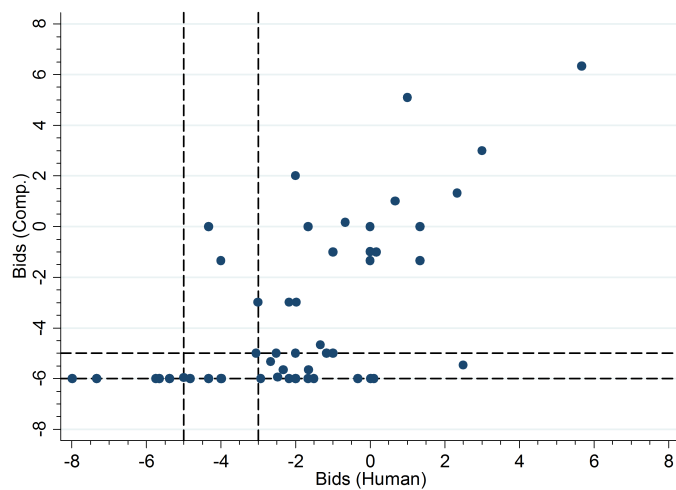


Figure A.1:  $\mathcal{AH}$  treatment - bid transition (Part I),  $N = 50$ .

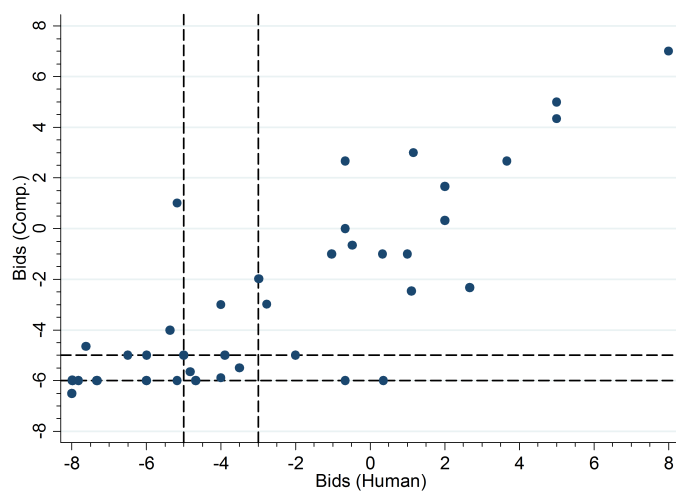


Figure A.2:  $\mathcal{AC}$  treatment - bid transition (Part I),  $N = 44$ .

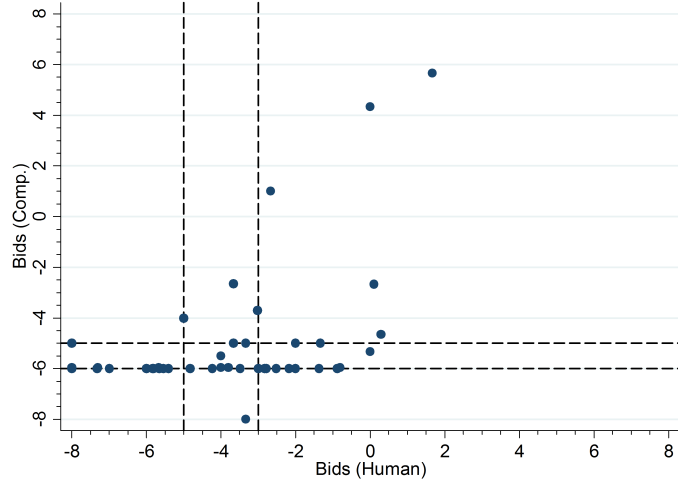


Figure A.3:  $\mathcal{TH}$  treatment - bid transition (Part I),  $N = 46$

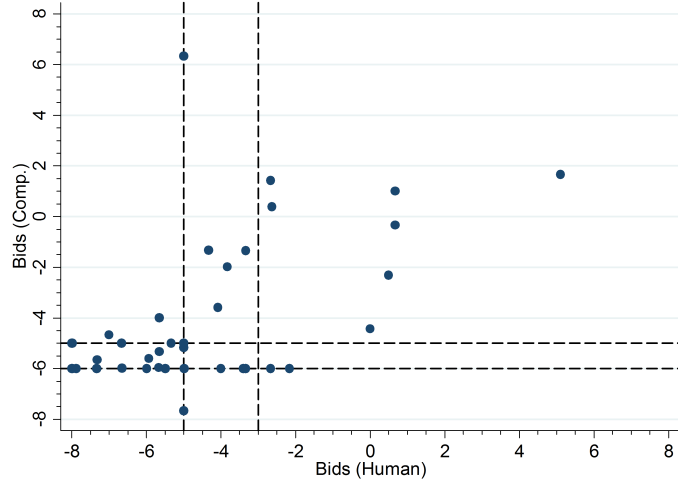


Figure A.4:  $\mathcal{TC}$  treatment - bid transition (Part I),  $N = 42$ .

## A.2 $\Delta\mu$ Calculation

| Adj. Av.                  | $\mu(\bar{b}(\cdot), b^e)$ | Adj. Av.                  | $\mu(\bar{b}(\cdot), b^e)$ |                                    |                                      |
|---------------------------|----------------------------|---------------------------|----------------------------|------------------------------------|--------------------------------------|
| $AH$                      |                            | $TH$                      |                            | $\Delta\mu(D_A; TH)$               |                                      |
| -1.80                     | 0.775                      | -4.00                     | 0.500                      | 0.275                              |                                      |
| $AH_{AC}$                 |                            | $TH_{TC}$                 |                            | $\Delta\mu(D_A; TH_{TC})$          |                                      |
| -2.62                     | 0.672                      | -4.64                     | 0.420                      | 0.253                              |                                      |
| $AC$                      |                            | $TC$                      |                            | $\Delta\mu(D_A; TC)$               |                                      |
| -3.97                     | 0.504                      | -5.56                     | 0.306                      | 0.198                              |                                      |
| $\Delta\mu(D_B; AH_{AC})$ |                            | $\Delta\mu(D_B; TH_{TC})$ |                            | $\Delta\mu(\{D_A, D_B\}; TH_{TC})$ | $\Delta^2\mu(\{D_A, D_B\}; TH_{TC})$ |
| 0.103                     |                            | 0.080                     |                            | 0.355                              | 0.023                                |
| $\Delta\mu(D_H; AC)$      |                            | $\Delta\mu(D_H; TC)$      |                            | $\Delta\mu(\{D_A, D_H\}; TC)$      | $\Delta^2\mu(\{D_A, D_H\}; TC)$      |
| 0.271                     |                            | 0.194                     |                            | 0.469                              | 0.077                                |

Table A.1:  $\Delta\mu(\cdot)$  and  $\Delta^2\mu(\cdot)$  calculated for Part I adjusted averages.

Table A.1 illustrates how  $\mu(\cdot)$ ,  $\Delta\mu(\cdot)$  and  $\Delta^2\mu(\cdot)$  can be calculated using the mean values of bids in different games. Bids in  $C$  games are adjusted as described in footnote 28. In the main text, we provide calculations for these measures using the regression analysis of table 5. The regression analysis provides a unifying framework that allows to quantify the impact of our cognitive complexities and their interaction using all available data and not only part I data as in table A.1. Moreover, in the regression the significance of different effects can also be easily verified.

## A.3 Proof of Proposition 1

In the auction game, particularly informative signals close to the boundary of the item value may influence the optimal strategy. Intuitively, these changes result from the information obtained about the value of the item. A signal within 3 units of the lower boundary implies that the true value is the higher one, thus increasing incentives to bid more for the item. However, the influence of this change of strategy through higher order beliefs vanishes quickly due to the discrete signals and their fixed distance of 6 in combination with the limited action space, as we will be outlined below in detail.<sup>40</sup> For signals  $x_i \in [46, 228]$ , bidding  $-8$  is optimal, as will be shown next. Notably, a signal within 3 units of the upper boundary implies that the true value can only be the lower one. There is thus no incentive to change the equilibrium strategy of bidding  $b = -8$ .

<sup>40</sup>In the original setup of Kagel and Levin (1986), signal noise follows a continuous, uniform distribution and actions are unconstrained. While the magnitude of the boundary effect quickly decreases, it never fully vanishes.



**Proof. *Equilibrium.*** We will first focus on signals  $x_i \in [46, 228]$  for which higher-order beliefs do not influence optimal behavior. To analyze optimal behavior in both the auction game and the transformed game, consider the best response function to the opponent's bid  $b_j$ . When abstracting from boundary signals and assuming that players bid type-independently, these best response functions describe the same optimal behavior for both games.<sup>41</sup> If player  $j$  bids high values,  $b_j \in [3, 8]$ , it is optimal for player  $i$  to never win the auction in either the auction game or the transformed game. The reason is that in this case, winning the auction would result in (weak) losses for sure in the auction game because the opponent is already bidding at least the commodity's value, even when she has received the lower signal. Analogously in the transformed game, the opponents is already bidding at least the higher value  $W_h^*$ . Hence, the best response is to bid anything that is relatively below the opponent's bid by at least 6 units,  $BR(b_j) \in [-8, b_j - 6]$ .

If player  $j$  bids values  $b_j \in [-8, 3]$ , it is optimal for player  $i$  to relatively underbid the opponent by slightly less than 6 points, making sure that he only wins the auction when he has received the higher signal or with 50% chance ("win in  $l$ " / "rule 3"). Hence, the best response function is  $BR(b_j) = b_j - 6 + \epsilon$ . By construction,  $b \in [-8, 8]$ , player  $j$  cannot bid low enough to cause a best response of overbidding by at least 6 points and thus always winning the auction ("win in  $\{l, h\}$ " / "rule 1"). Only if  $b_j \leq -15$  was possible, the best response would be  $BR(b_j) = b_j + 6$  since it would be more profitable to always win the auction than to only "win in  $l$ " / "rule 3".

With the best responses being either to underbid by at least 6 or by nearly 6, the unique equilibrium for both players is to bid  $b^e = b_i = b_j = -8$  in the transformed game. Similarly, any Nash equilibrium bid function takes the value  $b^* = -8$  for signals  $x \in [46, 228]$  (see discussion of boundaries below). Players then only win the auction when the lower item value realizes ("win in  $l$ " / "rule 3"), leading to an expected payoff of  $Eu_i = \frac{1}{2}(-3 - b_i) = 2.5$ . If player  $i$ , however, deviated to "win in  $\{l, h\}$ " / "rule 1", bidding  $b_i = b_j + 6 = -2$ , he would receive an expected payoff of only  $Eu_i = \frac{1}{2}(-3 - b_i) + \frac{1}{2}(+3 - b_i) = \frac{1}{2}(-1 + 5) = 2$ , showing that bidding  $-8$  is an equilibrium.

Additionally, any equilibrium bid function exclusively takes the value  $-8$  for signals  $x \in [46, 228]$  since at least one subject always have incentives to deviate from any pair of strategies in which not both subjects bid  $b_i = b_j = -8$ . When both players bid higher values than  $-8$ , at least one player has an incentive to underbid the other player because, as outlined before, best responses are either

---

<sup>41</sup>We show below that incentives for type-dependent bidding only exist near the boundary.

underbidding by at least 6 or nearly 6 (if such an underbidding is possible). These underbidding incentives only vanish when no underbidding is possible anymore and subjects bid  $-8$ . If only one player bids more than  $-8$ , this player has an incentive to also bid  $-8$  because of the outlined best response functions. These arguments also imply the uniqueness of the equilibrium in the transformed game.

*Boundaries.* In order to analyze whether the optimal strategies at the boundary influence strategies for interior signals in the auction game, we consider five player types. Player 5 receives a signal  $x^5 \in [46, 54)$ . His strategy might be influenced by his potential opponent with the lower signal: player 4, who receives the signal  $x^4 = x^5 - 6, x^4 \in [40, 46)$ . But player 4's strategy might of course be influenced by player 3 ( $x^3 = x^4 - 6, x^3 \in [34, 40)$ ) whose strategy might be influenced by player 2 ( $x^2 = x^3 - 6, x^2 \in [28, 34)$ ) and finally also by player 1 ( $x^1 = x^2 - 6, x^1 \in [22, 28)$ ).

Player 1 receives a signal  $x^1 \in [22, 28)$  from which he can infer that the commodity's real value is above his own signal. For this reason, player 1 cannot make any profits from bidding  $-8$ . Instead player 1 tries to overbid<sup>42</sup> player 2. But importantly, player 1 bids at most  $b^1 = +3$  because otherwise he would lose money because of overbidding the commodity's value  $x^1 + 3$ . Hence, in equilibrium, player 2 will bid  $b^2 \geq -3.01$  because any bid below would provide player 1 with an overbidding incentive that would lead player 2 to adjust his bid upwards. Additionally, player 2 cannot bid more than  $b^2 = 0$  because higher bids would lead to negative expected payoffs. Because of these incentives of player 2, in equilibrium, player 3 can ensure himself an expected payoff of at least  $Eu_i = 1.495$  by bidding  $b^3 = -5.99$ . If player 3 follows this strategy, player 2 cannot gain money by winning the auction, and, hence, player 2 will not overbid the player 3 and bids  $b^2 = -3$  to avoid losses. This, however, provides an incentive for player 3 to bid less than  $-5.99$ , which in turn provides an incentive for player 2 to overbid the third player and these overbidding incentives only fully vanish when player 3 bids  $-5.99$  again. Because of this circular incentive structure, in equilibrium, player 2 and player 3 will mix strategies. We do not fully characterize the exact mixed strategy equilibrium here, because it is sufficient for our purpose to show that players will not bid in certain intervals.<sup>43</sup>

As outlined before, for player 2, strategies above 0 cannot be part of an equilibrium. Hence player 3 can ensure himself a payoff of at least  $Eu_i = 1.495$  by

---

<sup>42</sup>More precisely, due to the rule we implement concerning equal bids, overbidding in this context means that player 1 only has to bid exactly player 2's absolute bid in order win the auction.

<sup>43</sup>The strategy space in our experiment is finite because participants have to round their bids to the cent-level. But for finite strategy spaces we know that there always exists an equilibrium.

bidding  $-5.99$ . Importantly, strategies that are part of a mixed strategy equilibrium must lead to a higher payoff than strategies that are not part of this equilibrium. Hence, bidding  $b^3 \in (-5.99, -2)$  cannot be part of a mixed strategy equilibrium because it leads to lower payoffs than bidding  $-5.99$ , independent of how player 2 exactly mixes pure strategies below  $b^2 = 0$ . Bidding  $b^3 \in [-8, -5.99)$  could in principle lead to the same payoff (or even a higher payoff) as bidding  $-5.99$  because the commodity's real value is underbid by a larger amount. The same is true for bidding  $b^3 \in [-2, -1.50]$  because player 3 might overbid player 4 with these bids. By bidding above  $-1.5$ , player 3 might still overbid player 4, but the (maximal) payoff ( $Eu_i = 1.49$ ) resulting from these bids is lower than the payoff of bidding  $-5.99$ . Bearing these considerations in mind, player 4 could always avoid to be overbid by player 3 by bidding  $b^4 = -7.49$  and ensuring himself a payoff of  $Eu^4 = 2.245$ . Because player 3, however, does not bid  $-5.99$  as a pure strategy but possibly also mixes strategies over  $[-8, -5.99]$  and  $[-2, -1.50]$ , player 4 potentially mixes strategies over  $-8 \leq b^4 \leq -7.49$ . Importantly, bidding above  $-7.49$  cannot be part of an equilibrium because then payoffs are lower than  $Eu^4 = 2.245$ . Especially overbidding player 5 even when this player is bidding  $b^5 = -8$  would only lead to an expected payoff of  $Eu^4 = 2.0$ . For this reason, the influence on strategies of boundary-signals ends at player 5: This player and all players with higher signals than player 5 will bid  $-8$  since their lower-signal opponents do not have an incentive to overbid them. In other words, for any signal  $x \in [46, 228]$ , any Nash equilibrium relative bid function takes the value  $b^*(x) = -8$ .

Additionally, at the higher boundary of the commodity's value space, no problems occur: A player receiving the signal  $x^{high} \in (222, 228]$  knows that the commodity's real value is below his own signal. Hence, he has to underbid his opponent who has a lower signal in order to earn money. But this does not lead to a change in equilibrium bid function because in case the opponent bids  $-8$ , the player with  $x^{high}$  also just bids  $-8$  and has no incentive to deviate. ■

## A.4 Communication data

In order to gather further, complementary evidence on belief formation and conditional reasoning, we implement the auction game with a communication design similar to Burchardi and Penczynski (2014), employing an additional 41 subjects. In this setting, teams of two players communicate about the bidding decision. In particular, the team partners first individually suggest a bid and can write a

justifying message. After this information is simultaneously exchanged, both team partners indicate – again individually – a final bid in a second step, of which one is randomly taken as the team’s bid. This randomly chosen bid – in conjunction with the other team’s bid – then determines both team members’ payoffs. Under the reasonable assumption that the best way to convince one’s team partner is by explaining to him the own reasoning, the design gives an incentive to write down the own reasoning process as fully and clearly as possible.

In this communication treatment, subjects play the first six periods of the  $\mathcal{AH}$  treatment: three periods of the auction game against humans, denoted as  $AH^{com}$ , and then three periods against computers,  $AC_{\mathcal{AH}}^{com}$ . Subjects enter six bids and messages before any communication is exchanged. This ensures that the bids and the messages exclusively reflect the subject’s individual reasoning. Our focus is on analyzing these messages – as an expression of subjects’ reasoning – but we note that the mean suggested bids are  $-1.24$  ( $AH^{com}$ ) and  $-1.73$  ( $AC_{\mathcal{AH}}^{com}$ ), comparable to our main data ( $-1.80$  in  $AH$  and  $-2.62$  in  $AC_{\mathcal{AH}}$ ).<sup>44</sup>

Based on the classification instructions reprinted in appendix B.6, a research assistant classified the message content according to various criteria. In contrast to common practice, we did not employ further research assistants as table A.3 reprints *all* individual, translated messages sent in  $AH^{com}$ . Thus, readers have the chance to fully examine messages and classification by themselves. For each criterion, table A.2 indicates the fraction of messages satisfying it and gives the mean suggested bid of the players whose message do not (0) and do (1) satisfy the criterion.

The table shows that 56% of subjects’ acknowledge the two possible states of the world (criterion Sa).<sup>45</sup> More specifically, 34% describe the two possible information signals of the other team (Sb) and 24% mention that bid ranges differ between states for the other team (Sc). Other state specific particularities beyond Sa are discussed by 34% of subjects (Sd). Messages with content according to each of these criteria are associated with significantly lower mean bids compared to those without. This shows that the deliberation of the possible states of the world allows subjects to bid closer to equilibrium. Notably, more specific deliberations (Sb, Sc, Sd instead of Sa) lead to even lower bids but fewer subjects engage in

<sup>44</sup>The slight differences are partially due to the first *com* session, which featured a previous version of the instructions and non-integer signals and values. The main results hold when excluding this session and mean bids are  $-1.31$  ( $AH^{com}$ ) and  $-2.35$  ( $AC_{\mathcal{AH}}^{com}$ ). Remaining differences might be due to the additional complexity of the communication setup or due to noise.

<sup>45</sup>Since subjects make three bids in  $AH^{com}$ , we aggregate subjects’ classification: As long as a criteria is fulfilled in one of the three messages, a subject fulfills this criteria (with the exception of I and Empty).

them, providing further evidence the need to condition provides an obstacle to subjects.

Only 15% of subjects reflect belief formation by deliberating the other team’s decisions (Ba). In that deliberation, 5% rule out particular actions (Bb) while 10% indicate likely choices (Bc). Despite the small number of messages satisfying these criteria, the significant differences in mean bids show that belief formation leads to bids much closer to equilibrium (Ba and Bb).

Looking at those 6 out of 41 (15%) subjects in detail does not reveal any evidence that subjects have particularly high beliefs that would rationalize weakly dominated play. It appears, however, that the beliefs of the six subjects are more complex than the degenerated one induced by the computer. Notably, when these beliefs are fairly concrete (ID 15: “nobody bids more than their information”) – as arguably in the case of three subjects (IDs 15, 19, 26) – subjects’ best response ability seems not to be overstrained as those subjects clearly make good decisions with average bids of  $-6.22$ . When these beliefs are fairly vague (ID 17: “all other teams enter smaller bids”) as in the case of IDs 9, 10, 17, subjects do not play very successfully due to average bids of  $-1.29$ , providing some indication that complex beliefs can sometimes overstrain best response abilities. Nonetheless, the more central message seems to be that 35 out of 41 appear not to form beliefs.

In line with this idea, different types of other deliberations – that are often less sophisticated and do not necessarily require forming beliefs – play a more prominent role. 22% communicate a tendency towards cautious bidding (Ra – ID 7: “bid not too high because we risk a loss otherwise”; ID 25: “#bettersafethansorry”) while 7% argue in favor of aggressive bidding (Rb – ID 40: “bid more”). Further, 10% of subjects include the unconditional expected value of the good in their argument (M – ID 12: “Try the mean value”). In addition, 20% of subjects only send messages that do not fulfill any criterion (I), often revealing either directly (ID 14: “I have no idea how we can decide”) or indirectly (ID 41: “What does the fuck say”) that their understanding of the game is unsatisfactory.

Since our ultimate interest is in behavior in auction games with human opponents, we focus on subjects’ deliberation in  $AH^{com}$  in this section. We note, however, that findings in  $AC_{\mathcal{AH}}^{com}$  are generally in line with more sophisticated behavior, as expected from the results with the main text. As examples, a majority of subjects, both describes the two possible information signals (56% Sb) and reflect belief formation (54% Ba).

| Criterion | Fraction | Mean rel. bid |       |       | $p$ -value |
|-----------|----------|---------------|-------|-------|------------|
|           |          | 0             | 1     | Diff. |            |
| Sa        | 0.56     | 0.13          | -2.33 | -2.46 | 0.005      |
| Sb        | 0.34     | 0.06          | -3.79 | -3.85 | 0.001      |
| Sc        | 0.24     | -0.52         | -3.52 | -3.00 | 0.007      |
| Sd        | 0.34     | -0.29         | -3.09 | -2.80 | 0.002      |
| Ba        | 0.15     | -0.82         | -3.75 | -2.93 | 0.053      |
| Bb        | 0.05     | -1.01         | -5.82 | -4.81 | 0.032      |
| Bc        | 0.10     | -1.09         | -2.72 | -1.63 | 0.450      |
| M         | 0.10     | -1.25         | -1.23 | 0.02  | 0.820      |
| Ra        | 0.22     | -0.75         | -3.01 | -2.26 | 0.048      |
| Rb        | 0.07     | -1.48         | 1.64  | 3.12  | 0.075      |
| I         | 0.20     | -1.70         | 0.63  | 2.33  | 0.056      |
| Empty     | 0.20     | -1.34         | -0.85 | 0.49  | 0.439      |

*Notes:* A subject is classified to fulfil one of the above criteria whenever any of the three messages fulfils the criterion. Only for “I” and “Empty”, all messages have to fulfil this criterion. The  $p$ -value results from a Wilcoxon ranksum test for equality of the mean bid between groups with messages that do not satisfy (0) and do satisfy (1) the criterion.

Table A.2: Fraction of subjects with satisfied messages criterion and group differences in mean bids ( $AH^{com}$ ,  $N = 41$ ).

Table A.3: Individual translated messages and belief classification.

| ID | Signal<br>$x_i$ | Bid<br>$b_i$ | Message  | Ba<br>message | Ba<br>subject |
|----|-----------------|--------------|--|---------------|---------------|
| 1  | 111.47          | -1.47        | Slightly risk-averse; maybe the others get a low signal; so less risk of loss, in case of prodt, this is higher; on the other hand the probability of winning is smaller, but I think it is worth it   |               |               |
| 1  | 155.91          | -3.66        | Bidding a bit under the mean value, so that the potential gain is higher and the loss smaller. However smaller chance of winning but also smaller chance of losses; Chance that the other team gets another range                                |               |               |
| 1  | 187.85          | -4.60        | In the lower end of the range: Higher profit and smaller loss; However a smaller chance of winning, but better to be +0 than to bid too much; Chance that the other team gets another range  |               |               |
| 2  | 108.05          | -1.66        | I would propose that we stay under our signal to be sure not to overbid, but also not too far away so that we still have the highest bid   |               |               |
| 2  | 156.40          | -1.47        | I think one should underbid the signal slightly so to have the highest bid and still potentially make a profit   |               |               |
| 2  | 195.76          | -1.64        |  |               |               |
| 3  | 102.35          | -2.35        | The odds that the price is higher are close to 8:4. I like this ratio. Loss is crap. I find it more reasonable to walk home with the endowment than to lose...   |               |               |
| 3  | 158.92          | -3.92        | We should not bid too much. Because 1. the possible gain is then too small or even negative and 2. we will get paid money even if we do not get the good   |               |               |
| 3  | 187.85          | -3.85        | We should not bid too much. Because 1. the possible gain in case we get the good is then small or negative, so that we should hope to get the good even with this low price and 2. we will get paid money paid even if we do not get the product |               |               |
| 4  | 112.35          | 2.65         |  |               |               |
| 4  | 156.18          | -1.18        | so far no special strategy   |               |               |
| 4  | 191.76          | -1.76        | see before   |               |               |
| 5  | 105.50          | -5.50        |  |               |               |
| 5  | 159.80          | -2.80        |  |               |               |
| 5  | 194.40          | -1.40        |  |               |               |
| 6  | 108.95          | -3.95        |  |               |               |
| 6  | 158.92          | -0.92        |  |               |               |
| 6  | 185.44          | 0.56         |  |               |               |

*Notes:* “Ba message” indicates the deliberation of the other team’s decision in the message and “Ba subject” recognizes the same deliberation in any message of the subject.

Table A.3: (continued)

| ID | Signal<br>$x_i$ | Bid<br>$b_i$ | Message   | Ba<br>message | Ba<br>subject |
|----|-----------------|--------------|---|---------------|---------------|
| 7  | 105.50          | -1.64        | My proposal is to not put the bid too high up. Roughly at 100 to 105.   |               |               |
| 7  | 155.91          | 3.86         | Purely random.  |               |               |
| 7  | 192.51          | -3.59        | Bid not too high because we risk a loss otherwise   |               |               |
| 8  | 102.35          | -1.35        | In any case a bit under the expected value  |               |               |
| 8  | 158.80          | -1.80        | In any case a little bit below the expected value.  |               |               |
| 8  | 195.76          | -1.76        | In any case a little bit below the expected value.  |               |               |
| 9  | 112.35          | -4.50        | Hey if we played alone, my proposal would be to always take $w^*$ . I think we should end up at 0 then. Since we play against other, I would reduce the bid by a little bit, maybe 3.50 EUR. If we don't get the good it doesn't matter anyways. ... Or just alternatively choose the minimal amount if 6 EUR are enough for you? It will probably not get much more... | ✓             | ✓             |
| 9  | 156.40          | -4.50        | Same thought.   |               |               |
| 9  | 185.44          | -3.94        | $w^*-2,94$  |               |               |
| 10 | 111.47          | 0.53         | I think the aothers will be rather more careful and bid less in the first round We should make use of that and take exactly the middle, so that we get the good, the loss should after all be bearable.   | ✓             |               |
| 10 | 158.80          | 1.20         | I would propose to get at it aggressively and make sure we get the product by bidding above our own price signal.   |               | ✓             |
| 10 | 194.40          | 1.60         | Tendency towards the higher value, so that we get the good Rather getting the good and have a chance to make a profit than not to be in because of a too small bid.   |               |               |
| 11 | 108.05          | 3.95         |   |               |               |
| 11 | 159.80          | 5.20         | $159 + (165-153/2)=165$   |               |               |
| 11 | 191.76          | 6.00         | The same Logic: $191,76 + (-185,76+197,76/2)=197,76$  |               |               |
| 12 | 108.95          | -0.95        | Try the mean value.   |               |               |
| 12 | 156.18          | 0.82         |   |               |               |
| 12 | 192.51          | 0.49         |   |               |               |
| 13 | 189.00          | -4.00        | 50:50 to be in the lower interval below $w^*$ and to be in the upper interval. If we are in the lower one we can go up until 192. If not it doesn't look so good.   |               |               |
| 13 | 140.00          | 0.00         |   |               |               |
| 13 | 142.00          | -3.00        |   |               |               |
| 14 | 189.00          | -4.00        | I have no idea how we can decide that   |               |               |
| 14 | 140.00          | 0.00         |   |               |               |
| 14 | 136.00          | 0.00         |   |               |               |

Notes: "Ba message" indicates the deliberation of the other team's decision in the message and "Ba subject" recognizes the same deliberation in any message of the subject.



Table A.3: (continued)

| ID | Signal<br>$x_i$ | Bid<br>$b_i$ | Message   | Ba<br>message | Ba<br>subject |
|----|-----------------|--------------|---|---------------|---------------|
| 15 | 183.00          | -6.00        |   |               |               |
| 15 | 146.00          | -6.00        | When the other team has received the +3 information, we win nothing, but a 50% chance to win 6 Euro since nobody bids more than their information will make, therefore always -6  | ✓             | ✓             |
| 15 | 142.00          | -6.00        |   |               |               |
| 16 | 183.00          | -5.00        |   |               |               |
| 16 | 146.00          | -1.00        |   |               |               |
| 16 | 142.00          | -1.00        |   |               |               |
| 17 | 183.00          | 0.00         | I think the value is 180. I am not fully in with the system to be honest. I am initially acting according to my feeling. Think both teams suspect to have gotten the higher signal  | ✓             |               |
| 17 | 146.00          | -2.00        | The chance to make losses is really high That is why all other teams enter smaller bids, that is why I orient myself rather far down If we are the same partners my proposal from round 1 was rubbish, I orient myself towards yours  | ✓             | ✓             |
| 17 | 136.00          | 0.00         | Pure gut feeling.   |               |               |
| 18 | 183.00          | -3.00        |   |               |               |
| 18 | 140.00          | -3.00        | No idea idea?   |               |               |
| 18 | 142.00          | -3.00        | No plan   |               |               |
| 19 | 183.00          | -7.00        | The team that gets the w*+3 information will always win the auction It only makes sense to bid 3 Euro less in order to avoid losses. So the other team with the lower information will also bid at least 3 Euro less than the value that they have obtained. If we have received the higher value, we can still win the auction with a bid of 8,90 Euro less and maximize our profit                      | ✓             |               |
| 19 | 146.00          | -7.00        | It only makes sense to bid less than the info value since the team with the higher value will win the auction To avoid losses, the team with the higher information value has to stay at least 3 Euro below the information value. Therefore also the other team will bid at least 3 Euro under ist value. It would be safe to bid 8.90 Euro under our information value, but only 7 Eur less is allowed. | ✓             | ✓             |

*Notes:* “Ba message” indicates the deliberation of the other team’s decision in the message and “Ba subject” recognizes the same deliberation in any message of the subject.

Table A.3: (continued)

| ID | Signal<br>$x_i$ | Bid<br>$b_i$ | Message  | Ba<br>message | Ba<br>subject |
|----|-----------------|--------------|--|---------------|---------------|
| 19 | 142.00          | -7.00        | The team with the higher information value will win the auction always. To avoid losses, it is rational for both teams to at least bid 3 Euro less than the respective information value. If we have the higher information value, the information for the other team is $142-6=136$ , and they should maximally bid 133 Euro. With the minimal amount of 135 Euro we would still win the auction and maximize our profit. | ✓             |               |
| 20 | 189.00          | -7.00        | Less profit results at a value of 186 Euro the other team would have the signal of 183 We could bid around 180? A little profit is better than none?   |               |               |
| 20 | 146.00          | -7.00        | If the value is 143 we should be relatively close  |               |               |
| 20 | 136.00          | -4.00        | If it is the smaller value, we at least have some profit. At the higher value even more. I would rather avoid the risk to suffer a loss.   |               |               |
| 21 | 189.00          | 2.25         |  |               |               |
| 21 | 146.00          | -0.77        |  |               |               |
| 21 | 142.00          | 1.02         |  |               |               |
| 22 | 183.00          | 3.00         |  |               |               |
| 22 | 140.00          | 5.00         | hopefully it is plus 3, lets see   |               |               |
| 22 | 136.00          | 4.00         |  |               |               |
| 23 | 188.00          | 0.00         | The item's value is either 185 or 191 Taler. If we choose the average as a bid, our potential gain and our potential loss will be the same. In my opinion it is justifiable risk   |               |               |
| 23 | 59.00           | 0.00         | The item's value is either 56 or 62 Taler. If we choose the average as a bid, our potential gain and our potential loss will be the same. In my opinion it is justifiable risk.  |               |               |
| 23 | 199.00          | -0.50        | The item's value is either 56 or 62 Taler. If we choose a bid slightly below the average, the gain is 3.5 and the loss is 2.5. In my opinion it is justifiable risk.   |               |               |
| 24 | 188.00          | 2.00         |  |               |               |
| 24 | 59.00           | 0.00         | I am sorry, my suggestion is just based on intuition and unfortunately not on any mathematical facts. I hope you have a better understanding   |               |               |
| 24 | 193.00          | 0.00         | We could change the price after the modification   |               |               |
| 25 | 188.00          | -2.99        | I would like to minimize the risk of losses and hope that we are shown the higher value. :) #bettersafethansorry   |               |               |
| 25 | 59.00           | -2.99        |  |               |               |
| 25 | 199.00          | -2.99        | I would like to minimize the risk of losses and hope that we are shown the lower value. :) #bettersafethansorry  |               |               |

Notes: "Ba message" indicates the deliberation of the other team's decision in the message and "Ba subject" recognizes the same deliberation in any message of the subject.

Table A.3: (continued)

| ID | Signal<br>$x_i$ | Bid<br>$b_i$ | Message  | Ba<br>message | Ba<br>subject |
|----|-----------------|--------------|--|---------------|---------------|
| 26 | 188.00          | -7.98        | I think it is smart to choose this number. If the item's value is 185, a price range of 179-185 will be displayed to the other team. If it plays safe, it should not be willing to bid more than 179.                                  | ✓             |               |
| 26 | 65.00           | -5.99        | If the value is 62, a price range of 56-62 will be displayed to the other team. They should not be willing to pay more than 59 Taler since they otherwise would make losses on average.  | ✓             | ✓             |
| 26 | 199.00          | -2.98        | If the value is 196, a price range of 193-199 will be displayed to the other team. They should not be willing to pay more than 196 Taler since they otherwise would make losses on average.  | ✓             |               |
| 27 | 188.00          | 8.00         | It always makes sense to bid the item's value +8. Thus, we get the item for sure   |               |               |
| 27 | 65.00           | 8.00         |  |               |               |
| 27 | 199.00          | 7.00         | it always makes sense to bid the highest possible item's value, thus bid $W^*+8$ to get the item for sure  |               |               |
| 28 | 182.00          | -2.00        |  |               |               |
| 28 | 65.00           | -4.00        |  |               |               |
| 28 | 193.00          | -5.00        |  |               |               |
| 29 | 182.00          | 3.00         | Shall we bid the higher one or ?   |               |               |
| 29 | 59.00           | 7.00         |  |               |               |
| 29 | 199.00          | 3.00         |  |               |               |
| 30 | 182.00          | 0.01         |  |               |               |
| 30 | 65.00           | 0.01         |  |               |               |
| 30 | 193.00          | 0.01         |  |               |               |
| 31 | 188.00          | -3.90        | I think, this is best to become the seller with a relatively high probability and at the same time definitely avoiding losses?   |               |               |
| 31 | 65.00           | -3.90        |  |               |               |
| 31 | 199.00          | -3.90        | I think, this is how to be safe, not risking to have a negative balance between the item's value and our price later, but at the same time making a profit   |               |               |
| 32 | 182.00          | -3.00        | If we bid less than 185 and it is below, we will make a loss, if we bid 179, we will make no gain, if we lose the auction no loss no gain or rather: 179 is safe, if the true value is 185 we will make a profit and otherwise no loss |               |               |
| 32 | 65.00           | -3.00        | safe no loss   |               |               |
| 32 | 193.00          | -3.00        | safe no loss   |               |               |

Notes: "Ba message" indicates the deliberation of the other team's decision in the message and "Ba subject" recognizes the same deliberation in any message of the subject.

Table A.3: (continued)

| ID | Signal<br>$x_i$ | Bid<br>$b_i$ | Message   | Ba<br>message | Ba<br>subject |
|----|-----------------|--------------|---|---------------|---------------|
| 33 | 182.00          | 3.10         |   |               |               |
| 33 | 59.00           | 3.10         |   |               |               |
| 33 | 193.00          | 2.10         | I try to bid as high as possible. Sorry, in case it does not work out :(  |               |               |
| 34 | 182.00          | -7.00        |   |               |               |
| 34 | 59.00           | -3.00        | If our bid is the highest, we will either get 0 or 6 Taler  |               |               |
| 34 | 193.00          | -3.00        | 0 or 6 Taler  |               |               |
| 35 | 188.00          | 3.00         |   |               |               |
| 35 | 59.00           | -2.00        |   |               |               |
| 35 | 193.00          | 5.00         |   |               |               |
| 36 | 182.00          | -5.00        |   |               |               |
| 36 | 59.00           | -5.00        | To make a profit, it is not only important to win the bidding but also to be below the item's value. Since our profit is determined by item's value-bid |               |               |
| 36 | 199.00          | -6.00        | I suggest those values with the highest gain  |               |               |
| 37 | 182.00          | -2.00        |   |               |               |
| 37 | 65.00           | -1.00        | 2 loss 4 gain   |               |               |
| 37 | 199.00          | -1.00        | 2 loss 4 gain   |               |               |
| 38 | 182.00          | 8.00         |   |               |               |
| 38 | 65.00           | -5.00        |   |               |               |
| 38 | 193.00          | 0.00         |   |               |               |
| 39 | 188.00          | -3.50        |   |               |               |
| 39 | 65.00           | -3.50        | With this bid we will definately not loose any money  |               |               |
| 39 | 193.00          | -3.50        | With this bid we will for sure not lose any money   |               |               |
| 40 | 182.00          | 3.00         |   |               |               |
| 40 | 59.00           | 6.00         |   |               |               |
| 40 | 199.00          | 3.00         | bid more!!  |               |               |
| 41 | 188.00          | -3.00        | done  |               |               |
| 41 | 59.00           | -3.00        |   |               |               |
| 41 | 199.00          | -3.00        | what does the fuck say  |               |               |

*Notes:* “Ba message” indicates the deliberation of the other team’s decision in the message and “Ba subject” recognizes the same deliberation in any message of the subject.

## B Online Appendix

### B.1 Behavioral Models

#### B.1.1 Cursed Equilibrium (CE)

CE models a limited ability to infer about types from actions. Therefore, in many games, a type-dependent action space prevents a meaningful application of CE (Eyster and Rabin, 2005). In the following, we propose to discuss it in our context by averaging strategies in the space of possible absolute bids. Due to the symmetry of our setup around a player’s signal, this approach should be true to the original idea of CE.

The prediction of full cursedness in our auction game is  $b_i^{CE} = -8$ , the same as in the transformed game and as the Nash equilibrium prediction. The intuition behind the equilibrium is the following: Player  $i$  with signal  $x_i$  faces an opponent with signal  $x_j = x_i - 6$  or  $x_j = x_i + 6$ . Being fully cursed, he thinks that *both* potential opponent types bid  $a_j = x_i - 6 + b_j$  or  $a_j = x_i + 6 + b_j$  with 50% each. Bidding  $b_i = -8$  in response to  $b_j = -8$  yields an expected payoff of  $Eu_i = 4$  (win in  $l$ ), whereas  $b_i = -1.99$  would only result in  $Eu_i = 1.99$  (win in  $\{l, h\}$ ).<sup>46</sup>

As outlined in the main text, conditional reasoning, as we understand it, requires two steps: (a) thinking in hypothetical situations and (b) conditioning on these hypothetical events when drawing appropriate conclusions on how to behave. In cursed equilibrium, agents remain able to think hypothetically, but draw misguided conclusions from these situations. As seen above, this alone is insufficient to explain overbidding behavior in our game.

Following CL, one has to assume a more general problem of conditional reasoning to explain behavior. In particular, subjects may not be able to think in hypothetical situations. More in line with our empirical results, a potential modeling approach could, hence, assume that subjects are not even able to distinguish between the two different hypothetical opponent types. Implicitly averaging bids for these types would then lead player  $i$  to believe that player  $j$ ’s bid is  $a_j = x_i + b_j$ . If we assume that player  $i$  processes the two possible item values in the same fashion and expects it to be  $x_i$ , the best response to  $b_j = -8$  is to overbid  $j$  with  $b_i = -7.99$ . These overbidding incentives only vanish when both players bid  $b = 0$ .

---

<sup>46</sup>We choose to describe the “win in  $\{l, h\}$ ” deviation to be  $b_i = -1.99$  since the response  $b_i = -2$  requires an understanding of the type-dependent tie-breaker. Moreover, the best response against the computer is the same with and without cursedness by a similar logic as outlined above.

### B.1.2 Level- $k$ Model

The level- $k$  prediction for players with positive level- $k$  that hold a uniform random level-0 belief is  $b^k = -8$ . The intuition is the following: Against a uniform random level-0 player, bidding  $b_i = -8$  implies winning only in state  $l$  with a probability of  $3/8$  and leads to an expected payoff of  $Eu_i = \frac{1}{2} \cdot \frac{3}{8} \cdot 5$ . Deviating to higher bids, however, leads to a reduction in the winning payoff that is not compensated by the increase of the winning probability. Since best response functions do not differ between the auction and the transformed game, there is no difference in the equilibrium across games.<sup>47</sup> It thus turns out that in our auction game, the type-dependence of the action space limits the interpretation of the level- $k$  model. In particular, level-1 players do not reflect informational naïveté. Otherwise, type-independent random level-0 play would lead to level-1 play that could be interpreted as informational naïve since it cannot extract information about signals (Crawford et al., 2013, p.28). In our case, by the definition of a best response, a level-1 player engages in conditional reasoning since the type determines the limits of the level-0 bid distribution.<sup>48</sup>

A model with a level-0 belief of truthful play,  $b^0 = 0$ , suffers from the same interpretative limitation. We can use it, however, as an exercise to look at the types emerging on the basis of such non-strategic and naïve behavior. Higher level players only differ in their beliefs. They all engage in conditional reasoning and, hence, do not suffer from inferential naïveté.

In particular, such a model predicts average bids  $b^0 = 0$ ,  $b^1 = -5.99$ , and  $b^k = -8$ , if  $k \geq 2$ , for both the auction and the transformed game. If we take noisy behavior into account in the simplest way and draw the line between types in the middle of their predicted bids, we get to predicted intervals as shown in table B.1. By treatment, the table further shows the number and fraction  $l^k$  of subjects falling into these categories for the games played against human opponents in part I.

In the presence of both conditioning and belief formation ( $AH$ ), the estimated fraction of level-0 players is very high, unlike many level- $k$  distributions previously estimated. Removing the need to condition and to form beliefs step by step leads to a normalization of the level- $k$  distribution to a point where it has a standard hump-shape and an average level of at least 1.02 ( $TC$ ).

---

<sup>47</sup>Against computerized opponents, since beliefs are fixed and level- $k$  maintains the best response assumption, no overbidding is predicted.

<sup>48</sup>We abstract from any influence the differences between the games could have on the level-0 belief.

Table B.1: Estimated level- $k$  distribution.

| Level- $k$   | $b^k$      | $AH$      |       | $TH$      |       | $AC$      |       | $TC$      |       |
|--------------|------------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
|              |            | $N$       | $l^k$ | $N$       | $l^k$ | $N$       | $l^k$ | $N$       | $l^k$ |
| 0            | $(-3, 8]$  | 36        | 0.72  | 16        | 0.35  | 21        | 0.48  | 9         | 0.21  |
| 1            | $(-7, -3]$ | 12        | 0.24  | 22        | 0.48  | 15        | 0.34  | 23        | 0.55  |
| $\geq 2$     | $[-8, -7]$ | 2         | 0.04  | 8         | 0.17  | 8         | 0.18  | 10        | 0.24  |
| <b>Total</b> |            | <b>50</b> |       | <b>46</b> |       | <b>44</b> |       | <b>42</b> |       |

Notes:  $\star C$  bids are adjusted (see footnote 28).

## B.2 Learning without feedback

In this section, we first provide evidence that our central results regarding conditioning and belief formation remain robust when considering single periods and not the average of the three periods per game, as done in the main text. Afterwards, we additionally analyze the data from part II of the  $AH$  and the  $TH$  treatment, as the main text only analyzes data from part I of our treatments.

### B.2.1 Single Period Play

In general, using the mean values for the three periods of each game leads to less noisy data than using single values. Moreover, the fact that subjects learn over time in some games but not in others should be interpreted as an additional result and not as weakness of our design. Nonetheless, we can explicitly incorporate this learning in our comparisons. As already outlined in the main text, subjects only improve their behavior in  $TH$  and  $TC$  (when these games are played first in  $TH$  and  $TC$ ). Thus, we only have to check comparisons involving these two games.

Regarding conditional reasoning, we want to analyze whether we observe a difference in  $\star H$  and  $\star C$  when we control for learning in  $TH$  and  $TC$ . When we compare bidding behavior and payoffs in  $\star H$  and this time base this comparison only on the first period, subjects still bid significantly less (and earn significantly more) in  $TH$  (Wilcoxon rank sum, bids -  $p = 0.018$ , payoffs -  $p = 0.050$ ). Additionally, plausible behavior is more likely in  $TH$  (Fisher's exact test,  $p = 0.011$ ) than in  $AH$ . When comparing behavior in  $\star C$  and considering only the first period, results still have the expected direction but are not generally significant (Wilcoxon rank sum, bids -  $p = 0.146$ , payoffs -  $p = 0.388$ ; Fisher's exact test,  $p = 0.057$ ).

Regarding belief formation, we want to analyze whether subjects still improve their behavior in  $TC$  compared to  $TH$  even if we incorporate that subjects learn

in the three periods of the  $TH$ . When we compare  $TH$  and  $TC$  and focus on third periods (to incorporate learning), the differences between the two settings naturally diminish and bids and payoffs are not significantly different any more (Wilcoxon rank sum, bids  $- p = 0.621$ , payoffs  $- p = 0.442$ ). Importantly, we have different equilibria in both settings. This biases against observing a difference in bids or payoffs. Hence, the more reliable measure to consider is whether the percentage of subjects playing plausible in both settings change. Indeed, more subjects play plausibly in the  $TC$  setting compared to  $TH$  and this difference remains highly significant (Fisher's exact test,  $p = 0.000$ ). Hence, as expected, results become slightly weaker when incorporating that subjects learn in the transformed games, but even then the overall pattern of the results remains intact.

Finally, we observe that learning in  $TH$  (when played first in  $\mathcal{TH}$ ) leads to a similar effect on subjects bids as playing this game after the computerized version,  $TH_{TC}$  ( $TH$  – third period bid:  $-4.68$  vs.  $TH_{TC}$  – mean bid:  $-4.64$ ). A possible explanation for this effect is that in  $TH$  subjects might use their first period bid as a first belief for the consecutive periods in similar fashion as the play against the computer provides a first belief.

### B.2.2 Part II Analysis

In the main text, our analysis focused on part I of the four treatments. In this section, we will additionally analyze part II of  $\mathcal{AH}$  and  $\mathcal{TH}$ . If problems with conditional reasoning are at the origin of the WC, we should observe a different learning pattern from part I to part II between the two treatments. If conditional reasoning is an obstacle for understanding the  $A$  games (both in  $H$  and  $C$ ), playing these game before the  $T$  games should not per se improve behavior in  $T$ . Subjects should not gain a better understanding of  $T$  via  $A$  simply because participants do not understand the  $A$  games because of the problems with conditional reasoning. Additionally, those subjects who manage to avoid the WC in  $A$  would most likely already play rationally in  $T$  if this game is played first. Playing  $T$  first, however, might very well facilitate playing  $A$ . By understanding the structure of  $T$ , a better understanding of the setting in which conditional reasoning on future events is necessary might arise. Hence, different patterns of learning behavior between the two treatments should be observed:

**Hypothesis 4:** *In  $\mathcal{AH}$ , no learning effect is observed in part II. Playing the  $T$  games after playing the  $A$  games leads to similar results as playing  $T$  first. In  $\mathcal{TH}$ , however, a learning effect is observed: Playing the  $A$*



Table B.2: Summary statistics -  $\mathcal{AH}$  &  $\mathcal{TH}$  treatments (Part I & Part II)

|             |         | <b>Part I</b>   |                 | <b>Part II</b>  |                 |
|-------------|---------|-----------------|-----------------|-----------------|-----------------|
| Mean        |         | $\mathcal{AH}$  | $\mathcal{TH}$  | $\mathcal{AH}$  | $\mathcal{TH}$  |
| (Std. dev.) |         | $A$             | $T$             | $T$             | $A$             |
| $H$         | Bids    | -1.80<br>(2.63) | -4.00<br>(2.61) | -3.66<br>(4.05) | -3.77<br>(2.88) |
|             | Payoffs | -0.56<br>(1.55) | 0.55<br>(1.37)  | 0.05<br>(2.29)  | 0.29<br>(1.90)  |
| $C$         | Bids    | -3.37<br>(3.30) | -5.00<br>(2.53) | -3.04<br>(3.74) | -4.48<br>(2.66) |
|             | Payoffs | 0.17<br>(1.53)  | 0.81<br>(1.56)  | -0.16<br>(2.09) | 0.68<br>(1.53)  |

*games after the  $T$  games leads to more plausible behavior than playing the  $A$  games first.*<sup>49</sup>

We focus on  $\mathcal{AH}$  and  $\mathcal{TH}$  because both treatments potentially provide a better comparison for the predicted learning effect than  $\mathcal{AC}$  and  $\mathcal{TC}$ . In the latter treatments, subjects also first play  $C$  in the second part which in principle could have an influence on  $H$ , which is at the very end of each treatment.<sup>50</sup>

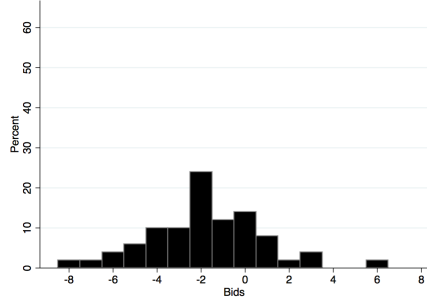
Table B.2 provides the mean values for subjects' bids and payoffs for part II of both treatments. Figure B.1 additionally shows histograms of subjects' bids in  $\mathcal{AH}$  and  $\mathcal{TH}$  for both parts.

When the auction game (against humans) is played after both  $T$  games,  $AH_{\mathcal{TH}}$ , only 28% of those subjects who win the auction face losses compared to 61% in  $AH$ . In line with this observation, bids in  $AH_{\mathcal{TH}}$  are lower than in  $AH$ , whereas payoffs are higher (Mean values - bids:  $-3.77$  vs.  $-1.80$ ; payoffs  $+0.29$  vs.  $-0.56$ ).<sup>51</sup> Hence, there is clear evidence that playing the  $T$  games in  $\mathcal{TH}$  before the  $A$  games helps subjects to avoid the WC in  $AH_{\mathcal{TH}}$ . Because of learning, we also do not

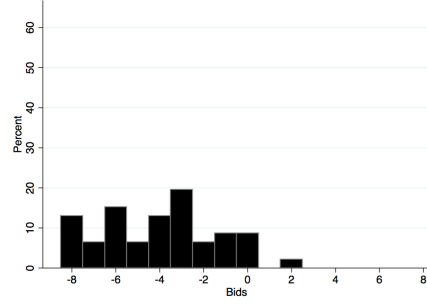
<sup>49</sup>Our design can, however, not distinguish whether such a learning effect is driven by the fact that subjects really understand the conditional reasoning because they played  $T$  first, or whether alternatively, subjects only understand that bidding low is a good strategy in  $T$  which they then also apply to  $A$ . It is, however, noteworthy, that subjects do not receive any monetary feedback before the end of the experiment.

<sup>50</sup>In general, results for  $\mathcal{AC}$  and  $\mathcal{TC}$  are comparable to those in  $\mathcal{AH}$  and  $\mathcal{TH}$  to the extent that playing  $A$  first does not help playing  $T$ , whereas playing  $T$  first helps playing  $A$  afterwards. This effect is significant in  $H$ , whereas in  $C$  the effect has the right sign but is insignificant. Hence, results in  $\mathcal{AC}$  and  $\mathcal{TC}$  are in general in line with our learning hypothesis. It might however, not be so clear, to what extent playing  $C$  first still influences these results.

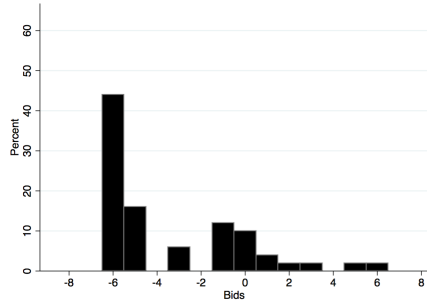
<sup>51</sup>Wilcoxon rank sum test - bids:  $p = 0.000$ ; payoffs:  $p = 0.002$ . Fisher's exact test based on plausible play -  $p$ -value = 0.025.



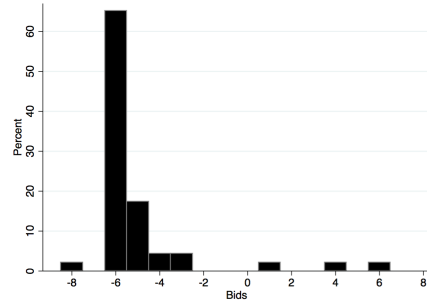
(a)  $AH$ ,  $N = 50$ .



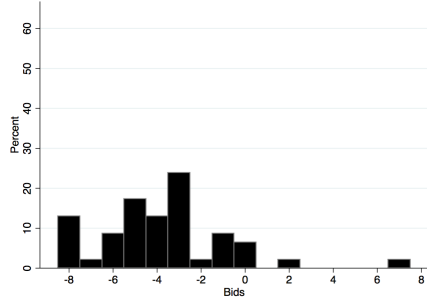
(b)  $TH$ ,  $N = 46$ .



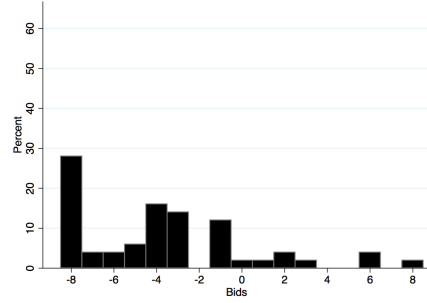
(c)  $AC_{AH}$ ,  $N = 50$ .



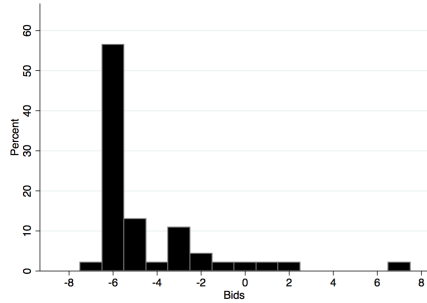
(d)  $TC_{TH}$ ,  $N = 46$ .



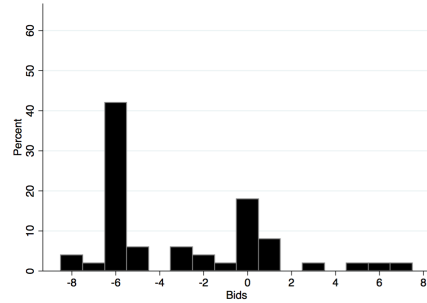
(e)  $AH_{TH}$ ,  $N = 46$ .



(f)  $TH_{AH}$ ,  $N = 50$ .



(g)  $AC_{TH}$ ,  $N = 46$ .



(h)  $TC_{AH}$ ,  $N = 50$ .

Figure B.1: Histograms of bids in  $\mathcal{AH}$  and  $\mathcal{TH}$ .

observe the treatment effect between the two games within-subject in the  $\mathcal{TH}$  treatment: Bids and payoffs are roughly the same between  $TH$  and  $AH_{\mathcal{TH}}$  (Mean values - bids:  $-4.00$  vs.  $-3.79$ ; payoffs:  $0.55$  vs.  $0.29$ ).<sup>52</sup>

Do we also observe this learning effect for  $C$ ? When the auction game is played after both  $T$  games,  $AC_{\mathcal{TH}}$ , only 22% of those subjects who win the game face losses compared to 45% in  $AC_{\mathcal{AH}}$ . In line with this observation, bids in  $AC_{\mathcal{TH}}$  are lower than in  $AC_{\mathcal{AH}}$ , whereas payoffs are higher (Mean values - bids:  $-4.48$  vs.  $-3.37$ ; payoffs  $+0.68$  vs.  $+0.17$ ), although not all differences are statistically significant.<sup>53</sup> Additionally, unlike in the case of  $H$ , the learning effect seems not to be strong enough to totally prevent a within-subject treatment effect.<sup>54</sup> Hence, there is some evidence for a learning effect in  $C$  of the  $\mathcal{TH}$  treatment, but this learning effect is weaker than in the  $H$  setting.

For the  $\mathcal{AH}$  treatment, we hypothesized above that subjects do not benefit from playing  $A$  before  $T$ . For  $TH_{\mathcal{AH}}$ , 47% of those subjects who win the game face losses compared to 32% in  $TH$ . Additionally, bids in  $TH_{\mathcal{AH}}$  are even slightly higher than in  $TH$ , whereas payoffs are lower (mean values - bids:  $-3.66$  vs.  $-4.00$ ; payoffs  $+0.05$  vs.  $+0.55$ ). Differences, however, are small and not statistical significant.<sup>55</sup>

Because subjects do not learn in the  $\mathcal{AH}$  treatment, we also observe the treatment effect between the two games within-subject in this treatment: Bids are higher in  $AH$  compared to  $TH_{\mathcal{AH}}$ , whereas payoffs are lower (mean values - bids:  $-1.80$  vs.  $-3.66$ ; payoffs:  $-0.56$  vs.  $+0.05$ )<sup>56</sup>

How does the behavior in  $C$  evolve in the  $\mathcal{AH}$  treatment? For  $TC_{\mathcal{AH}}$ , 43% of those subjects who win the game face losses compared to only 13% in  $TC_{\mathcal{TH}}$ . In line with this observation, bids in  $TC_{\mathcal{AH}}$  are higher than in  $TC_{\mathcal{TH}}$ , whereas payoffs are lower (Mean values - bids:  $-3.04$  vs.  $-5.00$ ; payoffs  $-0.16$  vs.  $+0.81$ ).<sup>57</sup> Hence, in the setting with computer opponents, we do not observe a learning effect, subjects in the  $\mathcal{AH}$  treatment perform even slightly worse than in the  $\mathcal{TH}$

---

<sup>52</sup>Wilcoxon signed rank test - bids:  $p = 0.814$ ; payoffs:  $p = 0.833$ . Additionally, a McNemar's test ( $p = 0.6072$ ) based on plausible play reveals no significant difference.

<sup>53</sup>Wilcoxon rank sum test - bids:  $p = 0.076$ ; payoffs:  $p = 0.054$ . But: Fisher's exact test based on plausible play:  $p = 0.301$ .

<sup>54</sup>Again, the statistical analysis is fairly inconclusive. A Wilcoxon signed rank test just reveals no significant difference (bids:  $p = 0.101$ ; payoffs:  $p = 0.371$ ) between  $TC_{\mathcal{TH}}$  and  $AC_{\mathcal{TH}}$ , but a McNemar's test based on plausible play reveals such a difference with marginal significance ( $p$ -value =  $0.065$ ).

<sup>55</sup>Wilcoxon rank sum test: Bids -  $p = 0.848$ ; payoffs -  $p = 0.293$ . Additionally, a Fisher's exact test based on plausible play supports this finding ( $p = 0.834$ ).

<sup>56</sup>Wilcoxon signed rank tests: bids -  $p = 0.000$ ; payoffs -  $p = 0.003$ . This result is also supported by a McNemar's test ( $p = 0.002$ ) based on plausible behavior.

<sup>57</sup>Wilcoxon rank sum test: bids -  $p = 0.033$ ; payoffs -  $p = 0.007$ . Fisher's exact test based on plausible play -  $p = 0.001$ .

treatment. For this reason, we also do not observe the treatment effect between the two games within-subject in the  $\mathcal{AH}$  treatment: Bids and payoffs are fairly similar in  $AC_{\mathcal{AH}}$  compared to  $TC_{\mathcal{AH}}$  (mean values - bids:  $-3.37$  vs.  $-3.04$ ; payoffs:  $+0.17$  vs.  $-0.16$ ).<sup>58</sup>

**Result 4:** In the  $H$  setting, we observe a learning effect as hypothesized: Playing the  $T$  games first facilitates playing the  $A$  games, whereas the reverse is not true. In  $C$ , a similar but weaker learning effect is observed in the  $\mathcal{TH}$  treatment. Overall, however, plausibility levels in the last game of both treatments are lower than expected. Exhaustion or increased confusion might be responsible for this result.

---

<sup>58</sup>Wilcoxon signed rank test: bids -  $p = 0.980$ ; payoffs -  $p = 0.205$ . McNemar's based on plausible play -  $p = 0.549$ .

### **B.3 Figures: Individual Data**

For completeness, figures B.2, B.3, B.4, and B.5 provide individual bids for all 12 periods of the experiment for all subjects of the four treatments. These figures support the evidence presented so far that subjects only improve their behavior in the transformed game when this game is played in part I of the experiment.

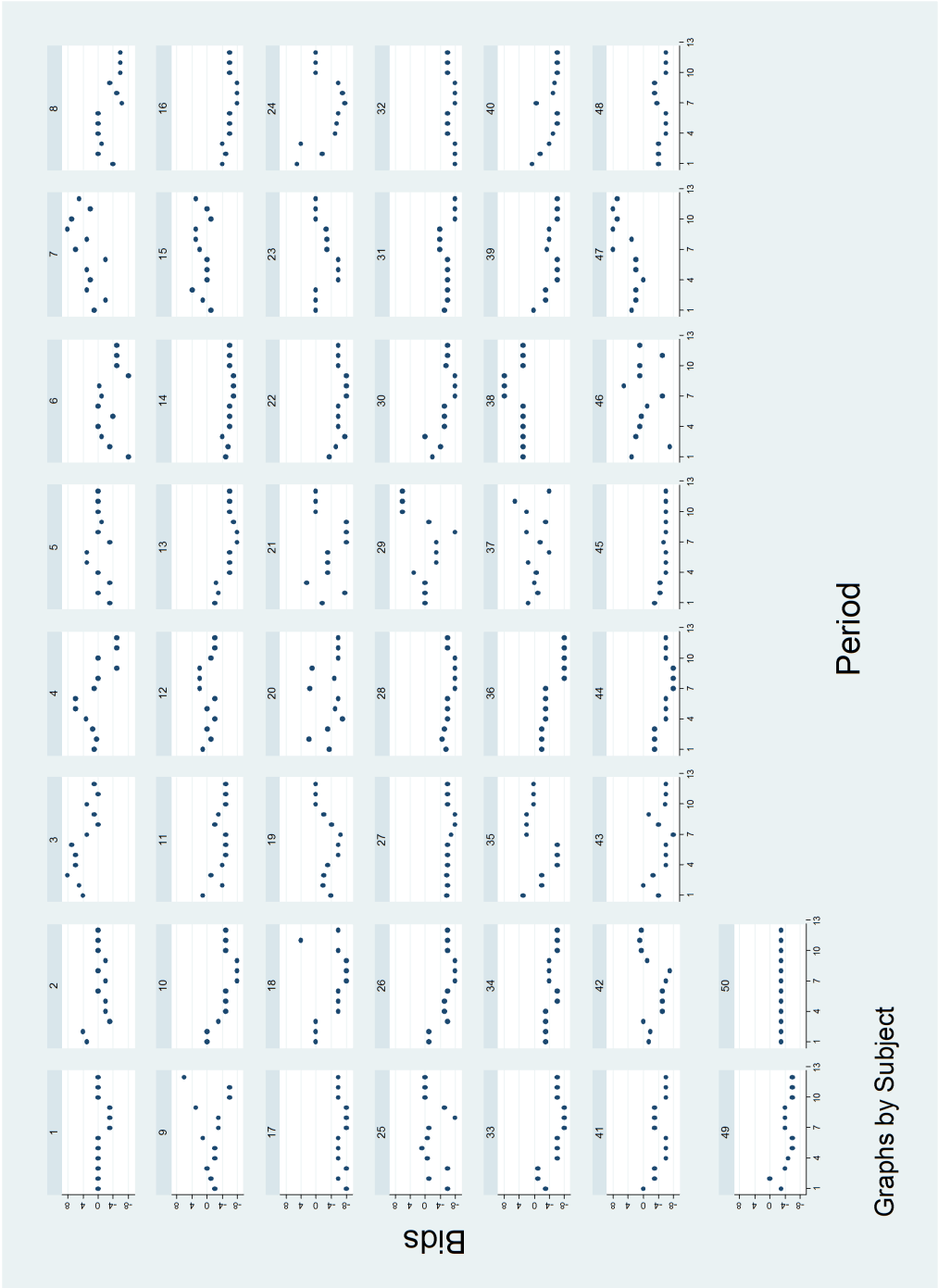


Figure B.2: Each subjects' behavior in the  $\mathcal{AH}$  treatment (3 periods per game)

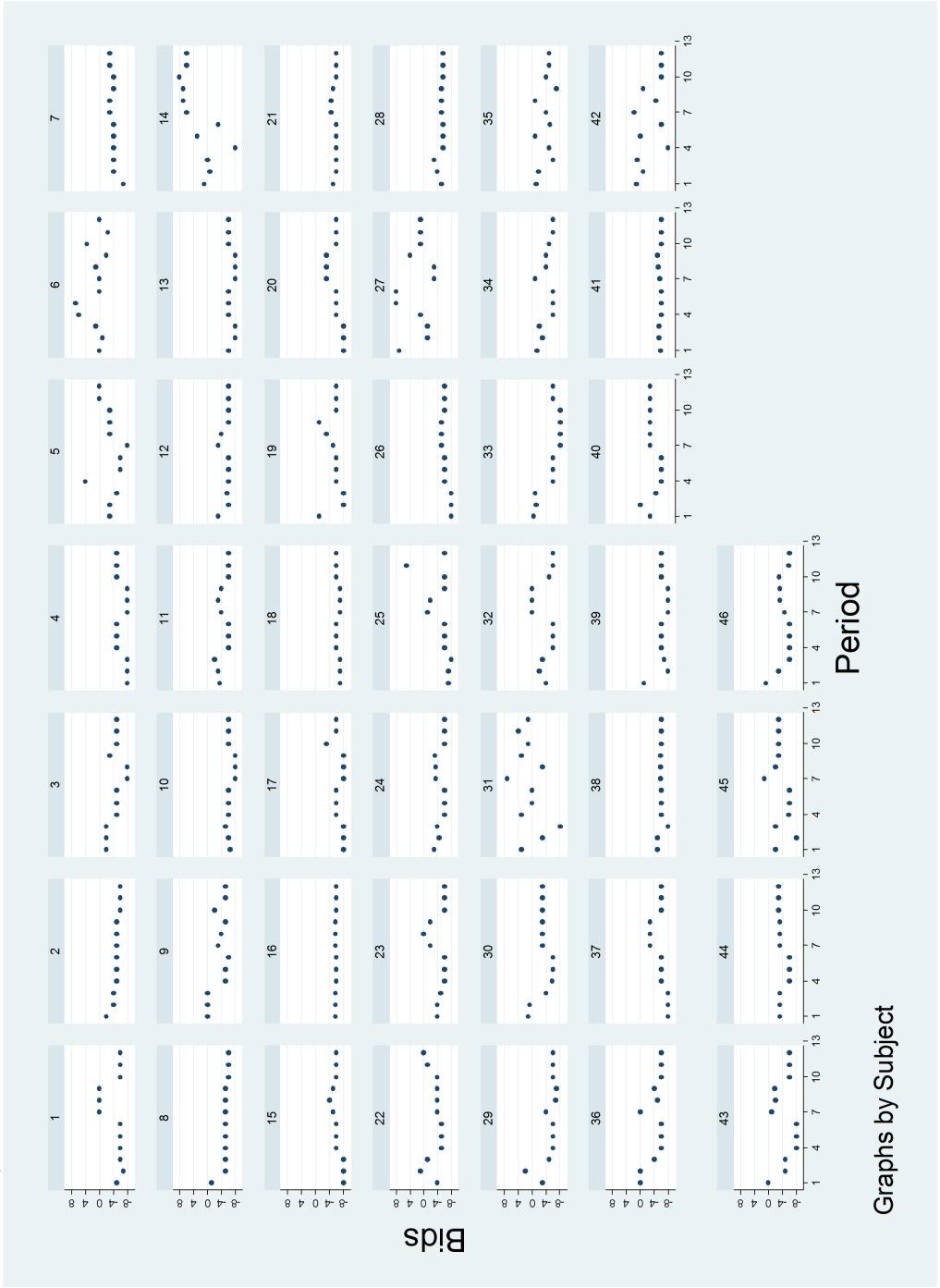


Figure B.3: Each subjects' behavior in the  $\mathcal{TH}$  treatment (3 periods per game)

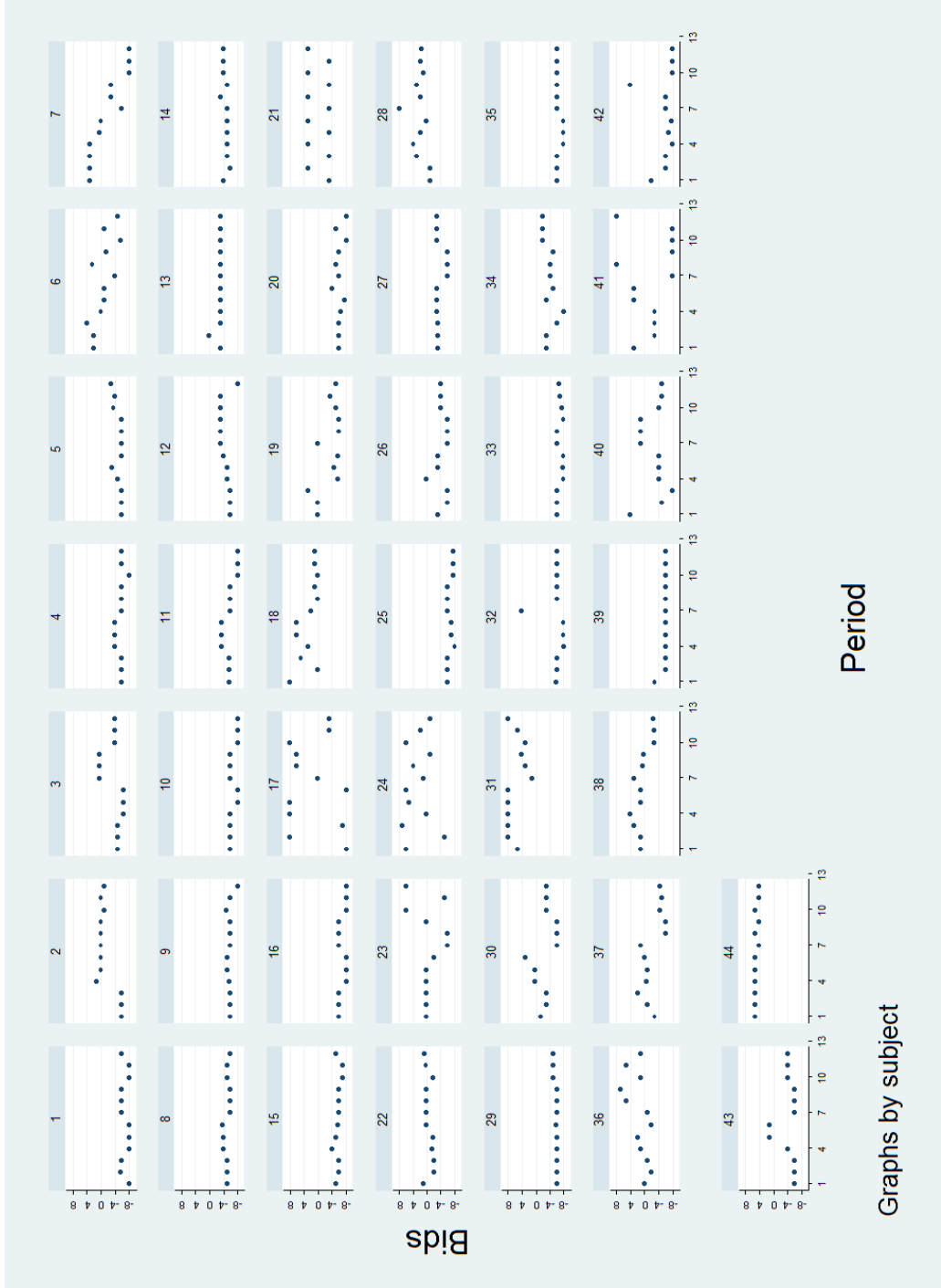


Figure B.4: Each subjects' behavior in the  $\mathcal{AC}$  treatment (3 periods per game)



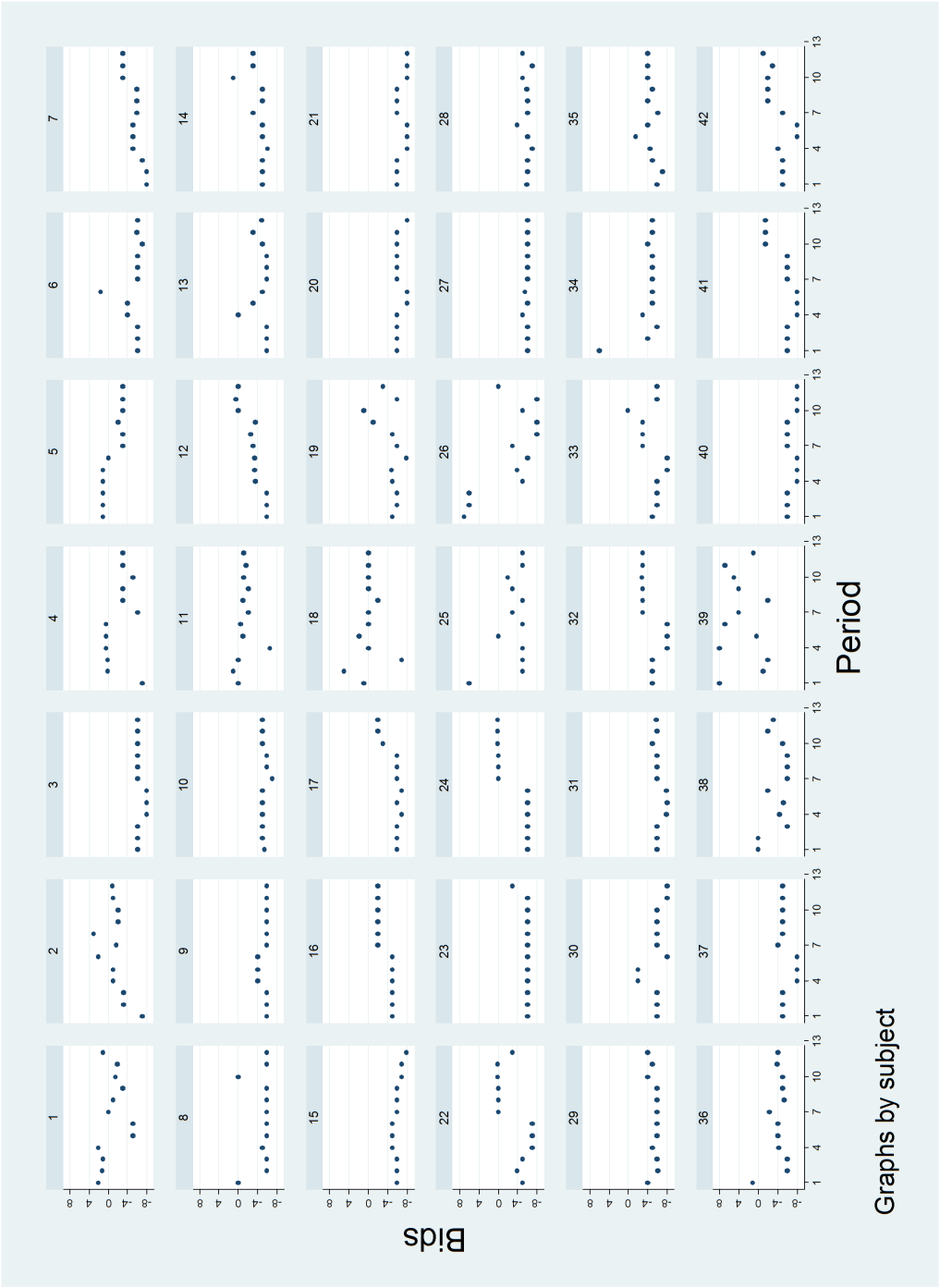


Figure B.5: Each subjects' behavior in the  $\mathcal{TC}$  treatment (3 periods per game)

## B.4 Instructions: $\mathcal{AH}$ treatment

**Welcome to the experiment!**

### Introduction

I welcome you to today's experiment. The experiment is funded by the University of Mannheim. Please follow the instructions carefully.

For participating, you first of all receive a participation fee of 4€. Additionally, you may earn a considerable amount of money. Your decisions and the decisions of other participants determine this additional amount. You will be instructed in detail how your earnings depend on your decisions and on the decisions of other participants. All that you earn is yours to keep, and will be paid to you in private, in cash, after today's session.

It is important to us that you remain silent and do not look at other people's screens. If you have any questions or need assistance of any kind, please raise your hand, and an experimenter will come to you. If you talk, shout out loud, etc., you will be asked to leave.

The experiment consists of three parts. For all three parts, you will receive separate instructions. You will first make your decisions for all three parts and only afterwards **at the very end** of the experiment get to know which payments resulted from your decisions. The currency used in all three parts of the experiment is called Taler. Naturally, however, you will be paid in Euro at the end of the experiment. **Two Taler will then convert to one Euro.**

If you have any questions at this point, please raise your hand.

### Part I

The first part of the experiment consists of  $2 \times 3$  trading periods (thus trading periods 1-3 and trading periods 4-6). These instructions describe the decision problem as it is present in trading periods 1-3. This decision problem will be slightly modified in the trading periods 4-6. You will be informed about the details of this modification at the end of trading periods 1-3.

In this part of the experiment, you will act as a buyer of a fictitious commodity. In each trading period, you will have the opportunity to submit a bid for one unit of the commodity. Importantly, not only you will have this opportunity to make a bid for the commodity. In each trading period, you will be matched with another

participant of this experiment. This participant will also have the opportunity to make a bid for the commodity. Importantly, you will always bid against another randomly determined participant in each trading period.

Your task is to submit bids for the commodity in competition with the other participant. The precise value of the commodity at the time you make your bids will be unknown to you. Instead, you and the other participant will receive an information signal as to the value of the item which you should find useful in determining your bid. Which kind of information you will receive, will be described below.

The value of the auctioned commodity ( $W^*$ ) will always be an integer and will be assigned randomly. This value can never be below 25 Taler and never be above 225 Taler.<sup>59</sup> Additionally, the commodity's value  $W^*$  is randomly and independently determined from trading period to trading period. As such a high  $W^*$  in one period tells you nothing about the likely value in the next period

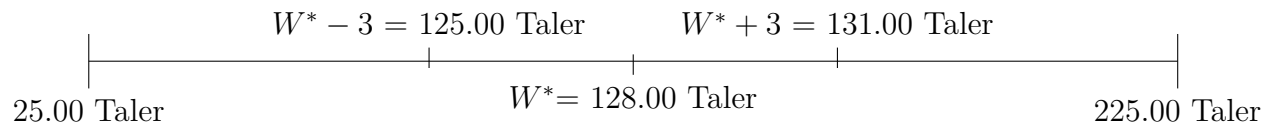
**Private Information Signals:** Although you do not know the precise value of the commodity, you and the participant who is matched with you will receive an information signal that will narrow down the range of possible values of the commodity. This information signal is either  $W^* - 3$  or  $W^* + 3$ , where both values are equally likely. In addition, it holds that when you receive the information signal  $W^* - 3$ , the person who is matched to you will receive the information signal  $W^* + 3$ . If in contrast, you receive the information signal  $W^* + 3$ , the other person gets the information signal  $W^* - 3$ .

For example, suppose that the value of the auctioned item (which is initially unknown to you) is 128.00 Taler. Then you will either receive a) the information signal  $W^* - 3 = 125.00$  Taler or b) the information signal  $W^* + 3 = 131.00$ . In both cases, the other person will receive the opposite information signal, in case of a) the information signal  $W^* + 3 = 131.00$  and in case of b) the information

---

<sup>59</sup>The instructions do not specify explicitly that the item value is *uniformly* distributed. This is, however, done implicitly by stating below that the item value is either three points above or below players' signals with *equal* probability. We followed this implementation to minimize the difference between the auction and the transformed game. The rules of the transformed game only state the two probabilities with which the lower or the higher item value realize but do not need to explain what a uniform distribution – a potentially problematic concept – is. We inform subjects about the uniform distribution after the example and do not discuss the unequal probabilities at the boundaries. Since no participant ever asked a question about this particular part of the instructions, participants must have inferred that probabilities naturally differ at the boundaries from the signal generation process shown in the example.

signal  $W^* - 3 = 125.00$  Taler. The line diagram below shows what's going on in this example.



It also holds that the commodity's value  $W^*$  is equal to the signal  $- 3$  or the signal  $+ 3$  with equal probability. The computer calculates this for you and notes it.

Your signal values are strictly private information and are not to be revealed to the other person. In addition, you will only be informed about the commodity's value  $W^*$  and the other participant's bid at the end of the whole experiment (when also the second and the third part of the experiment are completed).

It is important to note that no participant is allowed to bid less than the signal  $- 8$  and more than the signal  $+ 8$  for the commodity. Every bid between these values (including these values) is possible. Bids have at least to be rounded **to one cent**. Moreover, it holds that the participant who submits the higher bid gets the commodity and makes a profit equal to the differences between the value of the commodity and the the amount he or she bids. That is,

- Profit =  $W^*$  (128.00 Taler) – higher bid

for the higher bidding person. If this difference is negative, the winning person loses money. If you do not make the higher bid on the item, you will neither make a profit nor a loss. You will earn zero profits. If you and the other participant submit the same bid, the person who received the lower signal will get the commodity and he or she will be paid according to his or her bid.

At the beginning of part I, each individual participant will be given a starting capital credit balance of 8 Taler. Any profit earned by you in the experiment will be added to this sum. Any losses incurred will be subtracted from this sum. At the end of this part of the experiment, all gains and losses will be add up and the net balance of these transactions will be added to your capital credit balance. You are permitted to bid in excess of your capital credit balance. Even in case of a negative captial credit balance, you are still permitted to submit bids. Should your net balance at the end of this part of the experiment be zero (or less), you will not

get any payoff from this part of the experiment. But even in case you make losses in this part of the experiment, you will keep your initial show-up fee of 4€.

### **Summary:**

1. Two participants have the opportunity to submit bids for a fictitious commodity. The exact value of the commodity  $W^*$  is unknown to you. This value will, however, always be between 25 Taler and 225 Taler. Moreover, you receive a private information signal concerning the commodity's value. This signal is either  $W^* - 3$  or  $W^* + 3$ . The other participant will receive the other signal. No one is allowed to bid less than the signal  $- 8$  or more than the signal  $+ 8$ .
2. The higher-bidding participant gains the commodity and makes the following profit = commodity's value - higher bid.
3. Profits will be added to your initial capital starting balance. Losses will be subtracted from your initial capital starting balance. You can always submit higher bids than your capital starting balance.
4. This part of the experiment consists of two rounds with overall 6 trading periods. These instructions describe the decision problem as it occurs in the trading periods 1-3. There will be a modification of the decision problem for rounds 4-6, about which you will be informed soon.

If you have read everything, please click the “Ready” button, to start the experiment.

### **Modifciation of the decision problem**

You have now entered all decisions for the trading periods 1-3. Now, trading periods 4-6 will follow for which the decision problem so far will be slightly modified. As up to now the task is to submit bids for a fictitious commodity. Importantly, the other participant who also has the opportunity to submit bids will be replaced by the computer. As the other participant in the trading periods 1-3, the computer will also receive a signal about the commodity's value that is opposite to your own signal. The computer then decides according to the following

decision rule: ***The computer always exactly bids his information signal.*** Suppose, for example, that the true value of the commodity is 128.00 Taler. If the computer receives the information signal 125.00 Taler (commodity's value  $- 3$ ), the computer's bid is equal to 125.00 Taler. If the computer receives the information signal 131.00 Taler (commodity's value  $+ 3$ ), the computer's bid is equal to 131.00 Taler. Otherwise, everything else does not change.

If you have read everything, please click the "Ready" button, to continue with the experiment.

## Part II

The second part of the experiment consists of 3 trading periods (trading periods 7-9). In this part of the experiment, you will again act as a buyer of a fictitious commodity. In each trading period, you will have the opportunity to submit a bid for one unit of the commodity. Importantly, not only you will have this opportunity to make a bid for the commodity. In each trading period, you will be matched with another participant of this experiment. This participant will also have the opportunity to make a bid for the commodity. Importantly, you will always bid against another randomly determined participant in each trading period.

Your task is to submit bids for the commodity in competition with the other participant. In general, the value of the auctioned commodity will always be an integer and will be randomly determined. This value can never be below 25 Taler and never be above 225 Taler. At the beginning of each period, you and the other participant will be informed about the commodity's value. Importantly, however, there is a slight uncertainty about the value of the commodity. This value can take two different specifications in every period. The commodity can either be worth  $W_1^*$  or  $W_2^*$ , where both values always differ by 6 Taler and  $W_1^*$  always indicates the lower value. Which of the two values really realizes depends on chance and your bid as well as the other participant's bid and will be explained to you in more detail below. Both your bid and the other participant's bid are not allowed to be lower than  $W_1^* - 5$  or higher than  $W_2^* + 5$ . Every bid between these values (including these values) is possible. Bids have at least to be rounded **to one cent**.

To make the rules of the auction understandable, they will be explained in detail with the help of an example. Suppose that at the beginning of one period,

you are informed that the commodity's value is either  $W_l^* = 107.00$  Taler or  $W_2^* = 113.00$  Taler. You and the other participant are not allowed to bid less than  $W_l^* - 5 = 102.00$  or more than  $W_2^* + 5 = 118.00$  Taler. Who gets the commodity depends on your bid and the other participant's bid. Three rules apply:

**1. Your bid is 6.00 Taler or more higher than the other participant's bid:**

In this case, you will get the commodity for sure. With a 50 percent chance each the commodity's value then is either  $W_l^*$  (107.00 Taler) or  $W_2^*$  (113.00 Taler). Hence, your profit is:

- Profit =  $W_1^*$  (107.00 Taler) – Your bid      or
- Profit =  $W_2^*$  (113.00 Taler) – Your bid

Both scenarios are equally likely and the computer will randomly choose which scenario occurs. If one of the differences is negative and this scenario occurs, you will make a loss. The other participant will be paid according to rule 2.

**2. Your bid is 6.00 Taler or more below the other participant's bid:**

In this case, you will not get the commodity in any case and your profit is zero. The other participant will be paid according to rule 1.

**3. Your bid is less than 6.00 Taler above or less than 6.00 Taler below the other participant's bid:**

In this case, either you or the other participant get the commodity with a 50 percent chance and the computer will make this decision. The commodity's value is in any case  $W_1^*$  (107.00 Taler). Hence, in case you get the commodity, your profit is:

- Profit =  $W_1^*$  (107.00 Taler) – Your bid

In this case, the other participant earns zero Taler. If on the contrary, you do not get the commodity, your profit is zero and the other participant's profit is:

- Profit =  $W_1^*$  (107.00 Taler) – His/her bid

In both cases, it holds for the person who gets the commodity that this person will make a loss if the difference is negative.

At the beginning of part II, each individual participant will be given a starting capital credit balance of 8 Taler. Any profit earned by you in the experiment will be added to this sum. Any losses incurred will be subtracted from this sum. At the end of this part of the experiment, all gains and losses will be add up and the net balance of these transactions will be added to your capital credit balance. You are permitted to bid in excess of your capital credit balance. Even in case of a negative captial credit balance, you are still permitted to submit bids. Should your net balance at the end of this part of the experiment be zero (or less), you will not get any payoff from this part of the experiment. But even in case you make losses in this part of the experiment, you will keep your initial show-up fee of 4€.

You will only be informed about the other participant's bid and which value of commodity actually has realized at the end of the whole experiment (when also the third part of the experiment is completed).

### **Summary:**

1. Two participants have the opportunity to submit bids for a fictitious commodity. The value of commodity will always be between 25 Taler and 225 Taler. Because of uncertainty, the commodity's value can take two specifications  $W_1^*$  and  $W_2^*$ , where the difference between both values is always 6 Taler. No one is allowed to bid less than  $W_1^* - 5$  and more than  $W_2^* + 5$ .
2. If one person bids at least 6.00 Taler more than the other person, this persons gets the commodity for sure and either makes the profit =  $W_1^* - \text{his/her bid}$  or the profit =  $W_2^* - \text{his/her bid}$ . If one person bids at least 6.00 Taler less than the the other person, this person does not get the commodity in any case and makes a profit of zero Taler. If the difference of the bids is less than 6.00 Taler, both participants get the commodity with a 50 percent chance and make the following profit =  $W_1^* - \text{his/her bid}$  in this case.
3. Profits will be added to your initial capital starting balance. Losses will be subtracted from your initial capital starting balance. You can always submit higher bids than your capital starting balance.
4. This part of the experiment consists of 3 trading periods.

If you have read everything, please click the “Ready” button, to continue with the experiment.



### Part III

The third part of the experiment consists of 3 trading periods (trading periods 10-12). These 3 trading periods are almost identical to the trading periods 7-9 of part II. In addition, your capital credit balance of the end of part II will be the starting capital credit balance of this part. Hence, the payoff you receive from part II and part III of the experiment will finally depend on the amount of the capital credit balance at the end of this part of the experiment. In part III of the experiment, the following modification of the decision problem of part II is implemented: As up to now the task is to submit bids for a fictitious commodity. Importantly, the other participant who also has the opportunity to submit bids will be replaced by the computer. As the other participant in the trading periods 7-9, the computer is informed about both possible values of the commodity. The computer then decides according to the following decision rule: ***The computer always exactly bids the mean value of both values of the commodity (hence  $\frac{W_1^* + W_2^*}{2}$  or  $W_1^* + 3 = W_2^* - 3$ ).*** Suppose, for example, that the true value of the commodity is either  $W_1^* = 107.00$  Taler or  $W_2^* = 113.00$  Taler. The computer will then bid 110.00 Taler ( $\frac{107+113}{2} = 107.00 + 3.00 = 113.00 - 3.00$ ). Otherwise, everything else does not change.

If you have read everything, please click the “Ready” button, to continue with the experiment.

## B.5 Instructions: Frequently Asked Questions

### Auction game

1. *When I make my decision about which bid to submit, what kind of specific information do I have? Do I know the true value of the commodity?*

You do not know the commodity's value  $W^*$ . When making your decision, you only know your private information signal. You also do not know whether you received the “high” or the “low” signal. You only receive one number. With a 50 percent chance, you have received the high signal and with a 50 percent chance you have received the low signal. All this also holds correspondingly for the other participant.

2. *On what does it depend whether I get the commodity and how much do I earn should this situation arise?*

The person who submits the higher bid gets the commodity. The profit then is:  $W^* - \text{higher bid}$ . If both bids are exactly the same (meaning bids are also the same on the cent-level), the person with the lower signal gets the commodity.

3. *Which values am I allowed to bid?*

You are allowed to under- and overbid your personal information signal by up to 8.00 Taler. In addition, it is important that you are not only allowed to bid integers. For example, you could also bid 30.45 Taler instead of 30 Taler.

### Transformed game

1. *When I make my decision about which bid to submit, what kind of specific information do I have? Do I know the true value of the commodity?*

When making your decision, you know about two possible specifications of the commodity's value:  $W_1^*$  and  $W_2^*$ . Which of these values actually realizes in the end depends on your decision, the other participant's decision and chance.

2. *On what does it depend whether I get the commodity and how much do I earn should this situation arise?*

If you at least bid 6.00 Taler more than the other person, you will get the commodity for sure. Your profit will then be  $W_1^* - \text{your bid}$  or  $W_2^* - \text{your bid}$ , with a 50 percent chance each. Conversely it holds, that if you bid at least 6.00 Taler less than the other person, you will not get the commodity and your profit will be zero. If the difference of the bids is smaller than 6.00 Taler, either you or the other participant gets the commodity with a 50 percent chance and the computer will make this decision randomly. If the computer chooses you as the winner, your profit will be  $W_1^* - \text{your bid}$ .

3. *Which values am I allowed to bid?*

You are allowed to underbid the lower value of the commodity  $W_1^*$  by up to 5.00 Taler and overbid the higher value of the commodity  $W_2^*$  by up to 5.00 Taler. In addition, it is important that you are not only allowed to bid integers. For example, you could also bid 30.45 Taler instead of 30 Taler.

## B.6 Classification Instructions: Intra-team communication treatment

*Our classification exercise is slightly more elaborate than the results in appendix A.4 suggest. First, it comprises three additional criteria (6, 7, 72) that we – due to space constraints – do not discuss in our analysis. Since these criteria require a fairly high degree of strategic and game sophistication, not too many subjects fulfill them (9%, 15%, 2%). In addition, subjects doing well in these criteria often also discuss beliefs or deliberate about the state of the world. In particular, whenever a subject fulfils either criteria 7 and 72, he or she also fulfills at least one belief criterion. Thus, our simplified analysis seems still to fully capture the essential deliberations of our subjects. Second, our coder also classified the sequence in which specific criteria were reflected in the messages to illuminate how subjects’ reasoning evolves. It is, however, beyond the scope this paper to analyze this.*

### Important terms and concepts

First we will give some definitions that should help you to understand the questions that we ask you in the classification section.

In the following we will describe the classification process for the analysis of the experiment. Please read this document and the instructions for the experiment entirely in order to get an overview and then start the classification based on the player’s sent message and action proposal.

Please read the messages of each player, taking into account his proposed action. Below you find detailed instructions on how to classify each player. It is important that you limit yourself to making inferences only from what can clearly be derived from the message stated, i.e. do not try to think about what the player *might* have thought.

***IMPORTANT: Please note only those classifications for which you are certain. Also, confine yourself to the content of the message and do not fill any gaps by yourself. For example, if the statement indicates to simply keep going in a previously defined manner, then the classification of such a message is not repeating the one of a previous message, but simply reflecting that the player wants to stick to the plan. That’s why nothing is to be indicated if the message is not containing play-relevant content.***

The coding consists of a set of elements that might be found in a message. These elements and their relationship between each other will be indicated in detail

below.

## Elements possibly reflected in a message

- 0** The message indicates that the proposal is simply the result of intuition or a gut feeling. Further deliberation does not exist and the message actively suggested or acknowledged that further deliberation was not done.

*Example: "Gut feeling."*

- 1a** The two possible states of the world in terms of the possible values of the product are indicated. Either explicitly or implicitly by drawing inferences that indicate knowledge about the two states.

*Examples: "So the item can take values 153 or 159." "By underbidding the signal with -1, we either gain 4 or loose 2."*

- 1b** The two possible information signals of the other team are indicated, explicitly or implicitly.

*Example: "They therefore see either signal 150 or 162, right?"*

- 1c** The two possible bid intervals are mentioned to be different, explicitly or implicitly.

*Example: "Note that if they have the lower signal, they cannot bid above 158 anyways."*

**Logic 1** Generally, it is sufficient to indicate one out of the three elements since 1b implies 1a and 1c implies 1b. Distinctions can be made if the sequence of the sentence is such that first 1a is indicated and then 1c, for example. See last point "Sequence" in this list.

- 2** The unconditional mean value or expected value of the item is mentioned, calculated or in any way alluded to.

*Example: "Well, in expectation we have an item worth 156." "Should we bid the mean value -1?"*

- 3** Beyond 1, a difference between the two states/information signals/intervals is being mentioned or described. Alternatively, one of the two states/./.. is mentioned or described specifically in detail.

*Example: "If we have the higher signal, the product is worth 153 only and we are likely to win."*

**4a** The relevance of the other team's decision is being acknowledged or indirectly alluded to.

*Example: "If we have the higher signal, the other team has to bid much more than their signal in order to win the item."*

**4b** The behavior of the other team is concretely deliberated. In particular, some actions of the other team are ruled out, for example due to dominance.

*Example: "It does not make sense to bid more than the signal-3, so they won't do that."*

**4c** The behavior of the other team is concretely deliberated. In particular, some actions of the other team are indicated to be likely choices. These actions could be described in fairly concrete terms or rather vaguely.

*Examples: "I am sure that they will just bid their information signal." "The others's bid range is [44,60]. I think they will bid fairly conservative."*

**Logic 4** While 4a indicates that the other team's decision is somehow alluded to, 4b and 4c require a prediction of behavior to be made, which could be negative (4b) or positive (4c). 4b and 4c imply 4a. Possible classifications are thus 4a, 4b, 4c, and 4b+4c.

**5a** A tendency towards cautious bidding is being communicated which appears to result from risk aversion, loss aversion or other preferences.

*Examples: "We should not bid too high as we run the risk to get the item at an excessive price." "Safe!!!"*

**5b** A tendency towards aggressive bidding is being communicated which appears to result from risk lovingness, auction fever or other preferences.

*Examples: "If we bid too low, we are not going to win this auction." "Let's bid high!"*

**Logic 5** 5a and 5b are mutually exclusive.

**6** The message communicates the bidding trade-off, namely that a higher bid on the one hand increases the probability of winning and on the other hand lowers the profit in the case of winning (and vice versa for a lower bid).

*Example: "We should bid a bit lower, while this reduces the chances of winning, if we win we will earn big."*

**7** A strategy that is judged useful is attributed to the other team and shapes the beliefs about them.

*Example: “We should clearly not bid too high, hence, we should not be worried that the other team bids high either.”*

**72** The attribution in 7 occurs more than once.

*Example: “We should clearly not bid too high, hence, we should not be worried that the other team bids high either. At that point, we might just play as low as possible. If they do that as well, there is no reason for us to change and one of the two teams will make a handsome profit.”*

**No indication** The attribution of none of those elements to a message implies that this message was either empty or did not convey any content that was related to the structure of the game.

Four more general points:

**Implicitness** In many comments we specifically instructed that the elements might only be implicitly mentioned. This is a general possibility and should make you indicate any element as soon as the message implicitly reflects it.

**Sequence** For a more detailed look into these elements, the sequence with which they appear might be informative. The entry in the Excel-Sheet will allow to reflect at which position in the message a given element was found. If one elements cannot be judged as before or after the other, this can easily be reflected by giving them the same position number in the sequence.

**Example** An example is given in the Excel-Sheet. While this gives an example of the classification method, it is not certain that our view on these messages is to be adopted like that. Please change the classification of the example if you disagree with our view on the message content.

**Computer** The criteria and examples given so far are phrased in terms that one team faces another team. As you know, teams, however, face a computer in periods 4-6. Please apply the criteria described so far accordingly. Note that the knowledge about the computer’s strategy makes 4b an irrelevant category while 4a and 4b could still be indicated, even though it simply reiterates what subjects know from the instructions. Communication against the computer can be fairly implicit. While the following first example – in which the subject proposes a best response – would require a “1b, 3, 4c” classification, the second example – in which this not the case – would require “no indication”, as no clear inferences can be drawn.

*Example: “If the computer gets the higher value, we can only make losses.  
If the computer gets the lower value, it will bid 6 Taler less than our signal.  
Thus, we have to bid -5.99”  
“We should overbid our signal by 0.01”*

## Entering data in Excel

It is very important to stick to the coding instructions provided here for otherwise the classification would be useless for future research.

Thus consider the following:

1. It is very important that you double check whether the first 3 columns are filled correctly, i.e. that you enter the data for the correct subject and round.
2. Any field should be empty or filled with a natural number, indicating the position number in the sequence, 1, 2, 3, etc.
3. The last column indicated “Comment” is for the case that you want to add something that is not covered in the instructions for this classification.
4. If you find interesting elements that occur frequently but that have not been picked up by us, feel free to add a new column and mark all messages that contain the element. You can then specify to us in an email what exactly this element is.

If you have any questions please do not hesitate to contact us.

Note that the number of characters refers to the German original sentence.



## Amendments

After the first classification, the following amendments were made orally to the instructions.

1. In the computer treatment, except for extreme exceptions, there is usually no difference between 1b and 4c. Indicating the two signals and deliberating the computer's actions is equivalent as follows from the experimental instructions.
2. The difference between 1c and 3 is that 1c is specifically referring to the mentioning of the difference in bid intervals, while 3 is a category that captures any mentioning above 1.
3. It is not true that 5a or 5b imply any of the three 1 categories. Although 1 reflects an appreciation of the risky structure of the setting, categories 5 only refer to notions of risk attitudes/preferences.
4. The numbers of the categories are purely for identification, they do not reflect a hierarchy. Any logical connection between categories is explicitly described.
5. Since the classification is supposed to reflect in a structured way the content of the message, 4c is not automatically ticked in the computer treatment.
6. In the computer treatments, please note an explicit intention to mimic the computer in the comments.
7. If the message is partly or fully consisting of a reference to arguments written in previous decisions, this can be noted by indicating category 8. Please do not tick any category for contents of the references messages.