Central tendency bias in belief elicitation*

Paolo Crosetto† Antonio Filippin‡ Peter Katuščák§ John Smith¶

April 15, 2020

Abstract

We conduct an experiment in which subjects participate in a first-price auction against an automaton that bids randomly in a given range. The subjects first place a bid in the auction. They are then given an incentivized elicitation of their beliefs of the opponent's bid. Despite having been told that the bid of the opponent is drawn from a uniform distribution, we find that a majority of subjects report beliefs that have a peak in the interior of the range. This result is robust across seven different experimental treatments. While not expected at the outset, these single-peaked beliefs have precedence in the experimental psychology judgments literature. Our results suggest that an elicitation of probability beliefs can result in responses that are more concentrated than the objectively known or induced truth. We provide indicative evidence that such individual belief reports can be rationalized by well-defined subjective beliefs that differ from the objective truth. Our findings offer an explanation for the conservatism and overprecision biases in Bayesian updating. Finally, our findings suggest that probabilistic forecasts of uncertain events might have less variance than the actual events.

Keywords: belief elicitation, quadratic scoring rule, overprecision, conservatism

JEL: C72, C91

1 Introduction

Often when experimenters describe a stochastic distribution to subjects, it is assumed that they accept the distribution as given.[1] We conduct an experiment in which subjects participate in a first-price auction against a randomly bidding opponent who uses a strategy known to the subject: the bid is chosen from the uniform

*We thank Carlos Alós-Ferrer, Roberto Barbera, Johanna Hertel, and Lisa Saal. John Smith thanks Biblioteca de Catalunya. This research was supported by Rutgers University Research Council Grant #18-AA-00143 and by funding from the Max Planck Institute of Economics, Jena, Germany.

†Corresponding Author, Univ. Grenoble Alpes, INRAE, CNRS, Grenoble INP, GAEL, 38000 Grenoble, France, paolo.crosetto@inrae.fr.

‡Department of Economics, University of Milan, via Conservatorio 7, 20122 Milan, Italy.

§RWTH Aachen University, School of Business and Economics, Templergraben 64, 52064 Aachen, Germany.

¶Rutgers University-Camden, Department of Economics, 311 North 5th Street, Camden, New Jersey, 08102, USA.

[1] One exception is Camerer and Weigelt (1988), who find that subjects can bring "homemade" priors to the experiment that are different from those specified in the experiment. We also note that Prospect Theory (Kahneman and Tversky, 1979) suggests that subjects can behave as if probabilities were different from those given by the experimenter. This relationship is characterized by the probability weighting function. Also see Cohen, Plonsky, and Erev (2020) for more on the possibility that subjects do not read, understand, or believe experimental instructions, in a context of choice under risk.

distribution on a given range. Subjects first place a bid in the auction. They are then given an incentivized elicitation of their beliefs of the opponent's bid. Despite having been told that their opponent's bid is drawn from the uniform distribution, we find that subjects tend to report beliefs that have a single peak in the interior of the range. While not expected at the outset, these single peaked beliefs have precedence in the judgments literature.

When subjects estimate physical quantities (length, weight, loudness, etc.), the judgments[2] often exhibit a bias toward the mean of the distribution of the stimuli (Hollingworth, 1910; Poulton, 1979). For instance, if subjects are tasked to make judgments of lengths of lines, lines longer than the mean tend to be underestimated and lines shorter than the mean tend to be overestimated. In other words, there is a tendency to judge physical quantities to be closer to the mean than they actually are. This effect is sometimes referred to as the *central tendency bias*.[3] Considered in aggregate, the reported distribution of judgments tends to be less variable than the true distribution of line lengths.

The central tendency bias has been found in various other judgment settings, for instance weight (Jones and Hunter, 1982), distance (Radvansky, Carlson-Radvansky, and Irwin, 1995), loudness (Algom and Marks, 1990), and temporal duration (Jazayeri and Shadlen, 2010). Huttenlocher, Hedges, and Vevea (2000) also find the central tendency bias in judgments of the fatness of computer-generated images of fish, the greyness of squares, and the lengths of lines.[4]

The central tendency bias is not commonly studied in economic settings. However, some authors report a bias toward the center of an ordered action space in games. For instance, Arad and Rubinstein (2012) and Arad and Penczynski (2018) find a tendency to overweight the central battlefields in Colonel Blotto games, despite there being no strategic advantage of doing so. Even more striking, in some settings players forego potential profits because of their central bias. Rubinstein, Tversky, and Heller (1997) report that hiders in

[2]Judgments of uncountable physical quantities are common in the psychology literature. See, for example, Duffy, Gussman, and Smith (2019) for an experiment where subjects estimate line lengths in a choice setting.

[3]It is also sometimes referred to as the "regression effect" (Stevens and Greenbaum, 1966) or the "contraction bias" (Jou et al., 2004). The representativeness heuristic (Kahneman and Frederick, 2002; Kahneman and Tversky, 1973) makes similar predictions. Finally, the extremeness aversion literature (Simonson and Tversky, 1992; Chernev, 2004; Neumann, Böckenholt, and Sinha, 2016) finds that subjects tend to avoid extreme options in choice settings.

[4]Huttenlocher el al. (2000) offer an explanation for this experimental regularity (also see Huttenlocher, Hedges, and Duncan, 1991). The authors propose a *category adjustment model*, which posits that subjects compensate for their imperfect memory and imperfect perception by employing information about the distribution of stimuli. We note that, according to the category adjustment model, subjects learn the distribution through experience. Huttenlocher et al. (2000) refer to this as a *Bayesian* model since the information about the distribution is employed to maximize the precision of the judgments. However, Duffy and Smith (2019) analyze data from a replication of Huttenlocher et al. (2000) and find that non-Bayesian explanations have more support in the data than the category adjustment model. Duffy and Smith (2018) analyze the data from Duffy, Huttenlocher, Hedges, and Crawford (2010), and come to a similar conclusion about the predictions of the category adjustment model. See also Crawford (2019) and Duffy and Smith (2020).

a hide-and-seek game have a bias toward the center of the action space in a way that is exploitable by the seekers.

We say that the elicitation of beliefs in our experiment exhibits the central tendency bias because there appears to be such a bias toward the mean of the distribution. In the domain of beliefs, Kareev, Arnon, and Horwitz-Zeliger (2002) find that subjects tend to perceive a stochastic distribution learned through sampling to be less variable than it truly is.[5] Theoretically, the authors explain this finding by arguing that subjects estimate the true variance by sample variance, which, if uncorrected for the degrees of freedom, is a downward-biased estimate of variance. We find that underestimation of true variance is present also when the distribution is learned by description rather than by sampling. Further, note that the theoretical justification described above does not apply in our case. Our finding therefore suggests that the central tendency bias in reporting of judgments in general and beliefs in particular might be driven by an intrinsic reporting bias toward the center of the support range.

We also go a step further and investigate where such bias might come from in our setting. We first rule out that it is due to a lack of understanding that the true underlying distribution is uniform. We also find that it is not an artefact of our experimental design, operating through incentive incompatibility of the belief elicitation procedure or through payoff hedging. Finally, we provide indicative evidence that such individual belief reports can be rationalized by well-defined subjective beliefs that differ from the objective truth. This view is also supported by a finding that, between subjects, beliefs are correlated with auction bids.

Apart from contributing to the literature documenting biases in reported judgments and beliefs, our finding, if corroborated by future research, has a significant implication for designing experiments. It implies that subjects do not necessarily accept the distribution as given by the experimenter, even when the distribution is as simple as the uniform distribution. Moreover, our findings offer another explanation for the conservatism and overprecision biases in Bayesian updating. To the extent that our results generalize to other distributions, our findings also suggest that probabilistic forecasts (rather than point forecasts) of events will have less variance than the true variance of these events. We hope to stimulate more empirical work on these issues.

[5] Also see Marchiori, Di Guida, and Erev (2015).

2 Experimental design

2.1 Overview

We conduct an experiment where subjects engage in a first-price auction. Subjects are told the distribution of their random-draw opponent's strategy: a uniform distribution on a range of possible integer bids. The subjects then place a bid in the auction. Subsequently, subjects are given an incentivized elicitation of their beliefs of the random-draw opponent's strategy and their beliefs of winning the auction. The random-draw opponent's bid is determined by a physical draw of a token at the end of the experiment. A total of 379 subjects participated in the experiment.

2.2 Belief elicitations

After submitting a bid in the auction, subjects report their beliefs of the distribution of their random-draw opponent's strategy. The strategy space of the random-draw opponent is divided into 5 bins of equal size. Subjects allocate probability weights into the bins. An automatic checker verifies that these amounts correctly sum to 100. Screenshots of the elicitation procedure are provided in Appendix D.[6] The maximum that subjects can earn on this task is 20 ECUs, where 1 ECU=€0.20.

It is well-known that eliciting beliefs from subjects, particularly the full distribution of beliefs, can be a difficult endeavor.[7] We elicit the distribution of beliefs of the opponent's strategy by using the quadratic scoring rule (QSR), apparently first suggested by Brier (1950). Presenting the payoff formula (or a simulator based on it) should, under standard assumptions, establish incentive compatibility under risk neutrality (see subsection 4.2 for different risk preferences). However, subjects might find it hard to infer the expected payoff dominance of truthful reporting from the formula. Indeed, there is now a growing empirical evidence, from both lab and field, that individuals often do not report their private type truthfully, even though the underlying mechanism is incentive compatible (see, for example, Hassidim, Marciano, Romm, and Shorrer, 2017). Although there could be several underlying explanations, difficulty in perceiving the incentive compatibility

[6]Figure D2 shows a treatment in which the random-draw opponent's strategy is the uniform distribution on $\{1, \ldots, 60\}$ and a subject attempts to report weights that correctly sum up to 100. Figure D3 shows a treatment in which random-draw opponent's strategy is the uniform distribution on $\{1, \ldots, 100\}$ and a subject attempts to report weights that do not sum up to 100.

[7]See Winkler and Murphy (1970), Savage (1971), Matheson and Winkler (1976), Allen (1987), Manski (2004), Karni (2009), Offerman, Sonnemans, Van de Kuilen, and Wakker (2009), Blanco, Engelmann, Koch, and Normann (2010), Armantier and Nicolas (2013), Hossain and Okui (2013), Schlag and van der Weele (2013), Andersen, Fountain, Harrison, and Rutström (2014), Schotter and Trevino (2014), Schlag, Tremewan, and van der Weele (2015), Schlag and van der Weele (2015), and Harrison, Martínez-Correa, Swarthout, and Ulm (2017).

is probably a primary reason. Simply providing the formula does not guarantee that subjects understand the incentive properties of the elicitation procedure.

We therefore opt for an alternative approach that attempts to communicate the incentive properties of the mechanism directly via an intuitive and easy-to-understand payoff determination description and a set of advices. In particular, subjects are told: "*You will be paid based on how closely your estimates match your opponent's bid. The exact formula (the so-called quadratic scoring rule) is complicated and the experimenters will be happy to explain it after the end of the experiment to those who are interested. However, in order to maximize your expected earnings from this procedure, you should report these likelihoods truthfully according to what you believe.*"

A clarification is in order here. The optimal response to the QSR depends on whether "your expected earnings" is taken to refer to subjective beliefs or, instead, to the objective distribution. Since we take care through the instructions and control questions (see subsection 4.1 below) to make sure that subjects hold a correct representation of the objective distribution, we would expect the subjective beliefs to be uniform, making the issue moot. That said, subjects might still wrongly believe that a single bid draw is not equally likely to fall into any bin. In that case, while the uniform response maximizes the objective expected payoff, truthful reporting does indeed maximize subjectively expected earnings. Since we are interested in eliciting subjective beliefs, we provide advice consisting of four statements on how to report that link "your expected earnings" to subjective beliefs:[8] (1) report higher probabilities in bins that you believe to be more likely in comparison to bins that you believe to be less likely; (2) report equal probabilities in bins that you believe to be equally likely; (3) do not concentrate the reported probability in one or two bins if you are not quite sure that the opponent's bid is in that bin (those bins); (4) do concentrate the reported probability in one or two bins if you feel confident that the opponent's bid is in that bin (those bins).

It is our view that we provide proper incentives for truthful subjective belief reporting directly, without these having to be inferred from the payoff formula.[9] We also prepared a sheet with the payoff formula to be shown to subjects who requested seeing it after the experiment.[10]

Subjects are also asked to report their beliefs of winning (and not winning) the auction given their bid. Given the known strategy of their opponent, there is a well-defined objective probability of winning the auction given the bid. Similar to the previous elicitation, there is an automatic checker that verifies that the

[8]We present an abbreviated form here. For the full wording, see "Question 1 stage" in Appendix C.

[9]We note that Qiu and Weitzel (2016) and Fairley, Parelman, Jones, and McKell (2019) also do not show their subjects a formal expression for the scoring rule.

[10]This information sheet is available on OSF at https://osf.io/fy3mt/. Only 2 out of 379 subjects requested this information.

weights correctly sum to 100.[11] These beliefs are also incentivized with the QSR. The maximum that subjects can earn on this task is 20 ECUs. Screenshots are provided in Appendix D.[12]

2.3 Auction treatments

Subjects place a bid against a random-draw-bidding opponent in one of seven different auction treatments, in a between-subject design. In every treatment, the payoff is an induced value known by the subject minus the bid in the event of winning the auction, and 0 otherwise. Ties are resolved in favor of subjects.

The *auction 100/100* treatment is our baseline.[13] In this treatment, the subject's value of the object is 100 ECUs. The random-draw opponent has a strategy of placing a bid drawn from the uniform distribution on $\{1, ..., 100\}$. Subjects select a bid from $\{1, ..., 100\}$. This treatment elicits bids using a visual representation of the strategy space and of the lottery induced by each choice (see Figure D6 in Appendix C for a screenshot).

The *auction 100/100 without visualization* treatment is identical to the baseline, except that auction bids are elicited simply by typing a response from $\{1, ..., 100\}$ rather than with the aid of the visual representation.

In the *BRET* treatment, we implement the Bomb Risk Elicitation Task (BRET, Crosetto and Filippin, 2013). In this treatment, subjects face a 10×10 matrix with 100 numbered boxes. Of these, 99 are empty, while one contains a time bomb programmed to "explode" at the end of the task, i.e., *after* choices have been made. The bomb has an equal probability to be in any of the 100 boxes. Subjects decide how many of the 100 boxes to collect in the increasing order of their numbers. If the collection does not contain the bomb, the payoff is equal to the number of collected boxes. It is 0 otherwise. The *BRET* treatment is isomorphic to the *auction 100/100* treatment, with the number of *uncollected* boxes corresponding to the bid, and the same visual representation as the baseline.

In the *auction 80/100* treatment, subjects' value of the object is 80 ECUs. The strategy of the computerized opponent and subjects' bidding space is identical to the baseline. The bid is elicited using the same visual representation as in the baseline, and includes the possibility of overbidding. The instructions contain a warning against bidding more than 80.

The *auction 60/100* treatment is identical to the *auction 80/100* treatment, with the exception that subjects' value of the object is 60 ECUs and the instructions contain a warning against bidding more than 60 (see Figure

[11] See "Question 2 stage" in Appendix C for complete instructions.

[12] See Figure D4 for a screenshot in which a subject attempts to report weights that correctly sum up to 100. See Figure D5 for a screenshot in which a subject attempts to report weights that do not sum up to 100.

[13] In the name of the treatments, the number before "/" refers to the subject's value of the object and the number after "/" to the upper bound of the bidding range of the automaton.

D7 in appendix C for a screenshot).

The *auction 60/60* treatment sets the subjects' value of the object to 60 ECUs. The computerized opponent bids uniformly on $\{1, ..., 60\}$. Subjects select a bid from $\{1, ..., 60\}$. Bids are elicited with a visualization analogous to the baseline, but with the size of the matrix reduced to 10×6 boxes that are numbered $\{1, ..., 60\}$.

The *auction 60/60 expand* treatment is identical to the *auction 60/60* treatment, with the exception that subjects select a bid from $\{1, ..., 100\}$ and the visualization accommodates this by presenting a 10×10 matrix. However, the difference from the baseline is that the auction winning probabilities are derived from the random-draw opponent bidding range being $\{1, ..., 60\}$ rather than $\{1, ..., 100\}$. The instructions contain a warning against bidding more than 60.

2.4 Experimental details and earnings

The timing of the experiment is as follows. First, subjects are required to pass an unincentivized test of their understanding of first-price auctions and uniform probability distributions.[14] Then, they place their auction bid under 1 of the 7 experimental treatments. The subjects then respond to an elicitation of their beliefs of the distribution of the random-bid opponent's strategy and their beliefs regarding the probability of winning the auction given their bid.

The auction values, bids and prices are expressed in ECUs, where 1 ECU=€0.20. Each belief elicitation question can earn as many as 20 ECUs. One of these two elicitations is randomly selected for payment. Subjects are paid their earnings in the auction, the randomly drawn belief elicitation and a €2.50 show-up fee.

The sessions were conducted in German in the laboratory of the Max Plank Institute for Economics in Jena, Germany.[15] They lasted approximately 30 minutes and subjects earned €8.50 on average. We note that this amount was above the hourly wage available to our student subjects at the time of the experiment.

[14] More on control questions in Section 4.1 below. See section C.2 in Appendix C for the complete list of questions that subjects have to answer correctly before being allowed to proceed with the experiment.

[15] Appendix C contains the English translation of the instructions given to the subjects. The original German wording is available from the corresponding author upon request.

Table 1: Mean weights within bins

Treatment	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Auction 100/100	11.92	21.92	30.38	23.72	12.07
Auction 100/100 w/o	11.68	21.39	28.98	23.05	14.90
BRET	18.16	25.87	23.82	18.45	13.69
Auction 80/100	16.53	26.09	28.72	18.06	10.59
Auction 60/100	15.85	25.77	28.08	18.95	11.35
Auction 60/60	12.25	19.03	29.68	23.59	15.44
Auction 60/60 expand	14.94	22.45	24.15	21.61	16.85
Pooled	14.33	23.02	27.65	21.27	13.73

Note: We list the means of the weights reported within each of the 5 bins for each treatment and pooled across all treatments.

3 Results

3.1 Summary statistics

We define a *response* to be the collection of probability weights allocated to the 5 bins. The response in the lowest bin is labeled *Bin 1*, next, *Bin 2*, and so on. The 5 bins in *auction 60/60* and *auction 60/60 expand* treatments refer to ranges $1-12$, $13-24$, $25-36$, $37-48$, $49-60$. The 5 bins in the remaining treatments refer to ranges $1-20$, $21-40$, $41-60$, $61-80$, $81-100$. Table 1 summarizes the means of the reported weights within each bin. Figure 1 gives an overview of the full dataset, with each point representing the weight allocated by a subject to the given bin and the boxplots summarizing the distribution. The central tendency is robust across treatments and immediately evident at the aggregate level.

Analyzing the data at the individual level, we first explore the extent of the general deviation from the uniform response. We define a response to be *non-uniform* if a distribution of weights other than $(20, 20, 20, 20, 20)$ is reported. Results are shown in Table 2. More than 72% of subjects give a response other than the uniform distribution. To capture how far the response is from the uniform distribution, we measure the largest vertical distance (d) of the cumulative distribution function (cdf) of the reported belief from the cdf of the uniform distribution at the four boundaries dividing the 5 bins. Cumulative beliefs of more than two thirds of subjects deviate from the uniform distribution by at least 10 percentage points, those of more than two fifths by at least 20 percentage points and those of more than a fifth by at least 30 percentage points.[16]

[16]Interestingly, not having the visualization (in *Auction 100/100 without visualization*) does not make the deviation from the uniform distribution larger. If anything, the deviation rate and the threshold distance are lower than the average of the other 6 treatments, but the difference is not statistically significant using the Fisher's exact test (the p-values corresponding to the four measures presented

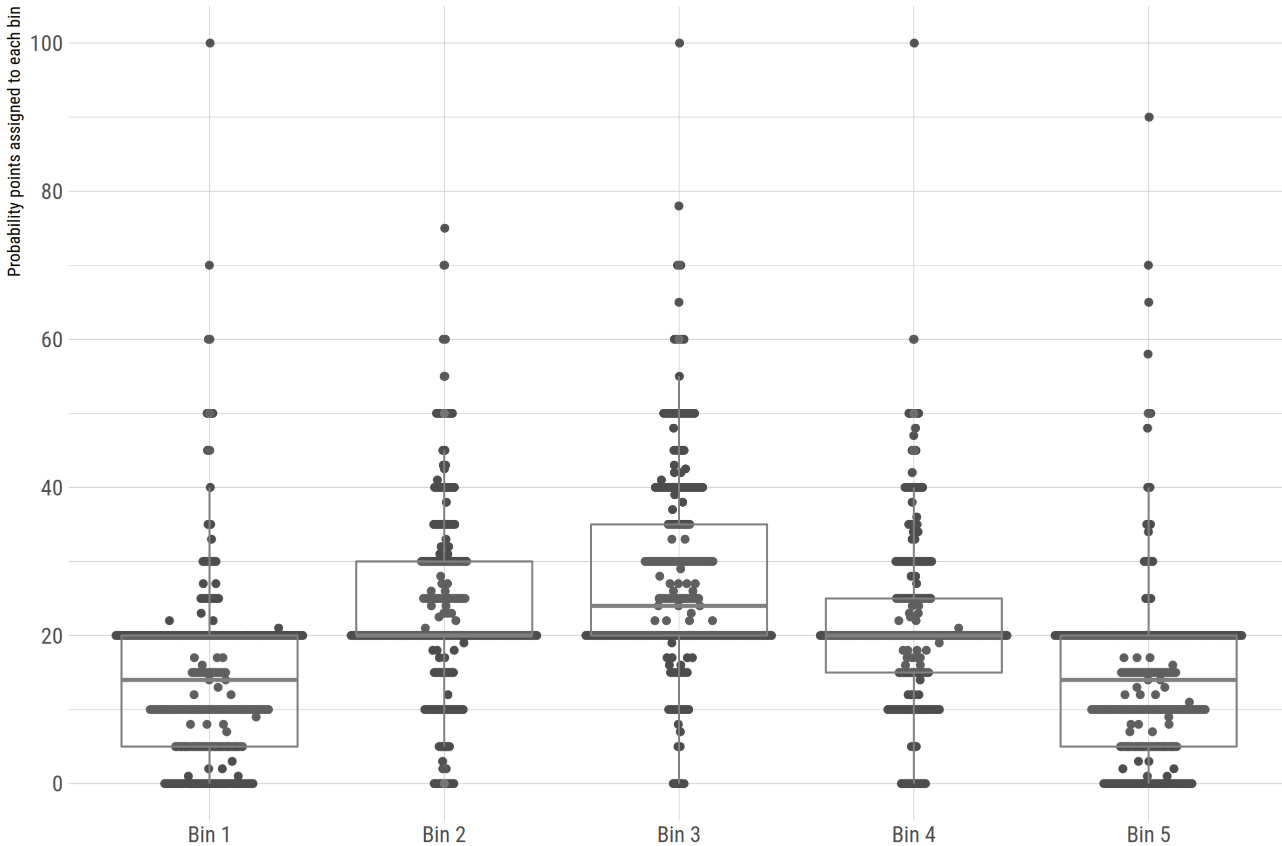

Figure 1: The full dataset at a glance. Each point represents the weight allocated by a subject to the given bin. The boxplots show the 25 th, 50th and 75th percentile of the distribution.

3.2 Central single-peaked responses

We next explore to what extent deviations from the uniform distribution are due to the central tendency bias. We define a response to have a *strict central single peak (strict-CSP)* if the weight in Bin 1 is strictly less than that in Bin 2, the weight in Bin 2 is strictly less than that in Bin 3, the weight in Bin 4 is strictly less than that in Bin 3, and the weight in Bin 5 is strictly less than that in Bin 4. If we define w_i as the weight allocated into Bin i, then we can write the definition of strict-CSP as $w_1 < w_2 < w_3 > w_4 > w_5$. To loosen this definition such that it allows a multi-bin peak that includes Bin 3, we define a response to have a *weak central single peak (weak-CSP)* if the inequalities are allowed to be weak, but the weight in Bin 1 is strictly less than the weight in

in Table 2 are 0.427, 0.880, 0.474 and 0.177, respectively).

Table 2: Non-uniform responses and distances from the uniform distribution

Treatment	Non-Uniform	$d > 0.1$	$d > 0.2$	$d > 0.3$	Subjects
Auction 100/100	46 (76.7%)	41 (68.3%)	20 (33.3%)	11 (18.3%)	60
Auction 100/100 w/o	40 (67.8%)	39 (66.1%)	22 (37.3%)	9 (15.3%)	59
BRET	44 (71.0%)	39 (62.9%)	20 (32.3%)	13 (21.0%)	62
Auction 80/100	25 (78.1%)	25 (78.1%)	18 (56.3%)	11 (34.4%)	32
Auction 60/100	40 (72.7%)	39 (70.9%)	28 (50.9%)	14 (25.5%)	55
Auction 60/60	41 (69.5%)	38 (64.4%)	29 (49.2%)	16 (27.1%)	59
Auction 60/60 expand	39 (75.0%)	35 (67.3%)	23 (44.2%)	10 (19.2%)	52
Pooled	275 (72.6%)	256 (67.5%)	160 (42.2%)	84 (22.2%)	379

Note: We list the number (and percentage) of subjects with non-uniform responses. We also list the number (and percentage) of subjects whose reported beliefs deviate from the uniform distribution by at least a threshold sup-norm distance $d \in \{0.1, 0.2, 0.3\}$.

Bin 3 and the weight in Bin 5 is strictly less than the weight in Bin 3.[17] To loosen the definition of strict-CSP in a different direction, namely by allowing a strict peak also in other non-boundary bins, we define a response to have a *strict semi-central single peak (strict-semi-CSP)* if it satisfies the conditions for strict-CSP or there is a strict single peak in either Bin 2 or Bin 4.[18] Finally, in the least restrictive definition, that subsumes both weak-CSP and strict-semi-CSP, we define a response to have a *weak semi-central single peak (weak-semi-CSP)* if it satisfies the conditions for weak-CSP or there is a weak single peak in either Bin 2 or Bin 4.[19] Table 3 describes the distribution of responses according to these definitions. Within each treatment and across all treatments, there appears to be a single peak in the interior of the bid support. Over 50% of responses satisfy our definition of a weak semi-central single peak. Figure 2 gives a visual characterization of the data by plotting the allocation of subjects to types of peaked response by treatment, assigning subjects to the most restrictive type.[20]

To further explore the incidence of a central single peak in the responses, we take pairwise tests of the differences of the weights between adjacent bins ($w_2 - w_1; w_3 - w_2; w_4 - w_3; w_5 - w_4$). We perform this analysis with paired t-tests and signed rank tests. Since we make multiple pairwise tests, we perform the Bonferroni correction. As we make four pairwise comparisons, we multiply every p-value by 4. We summarize this analysis in Table 4. We note that each statistic involving Bins 1 or 2 is positive and each statistic involving Bins 4

[17] In other words, a response has a weak-CSP if $w_1 \leq w_2 \leq w_3 \geq w_4 \geq w_5$, $w_1 < w_3$, and $w_3 > w_5$. Note that if a response has a strict-CSP, it has also a weak-CSP.

[18] A response has a strict-semi-CSP if it has a strict-CSP or it is a response with $w_1 < w_2 > w_3 > w_4 > w_5$ or $w_1 < w_2 < w_3 < w_4 > w_5$.

[19] A response has a weak-semi-CSP if it has a weak-CSP or it is a response with $w_1 < w_2 \geq w_3 \geq w_4 \geq w_5$ and $w_2 > w_5$, or it is a response with $w_1 \leq w_2 \leq w_3 \leq w_4 > w_5$ and $w_1 < w_4$. Note that if a response has a strict-semi-CSP, it has also a weak-semi-CSP.

[20] For instance, if a response has a strict-CSP, then, by definition, it has also a weak-CSP; in this case, the subject is assigned to the strict-CSP type.

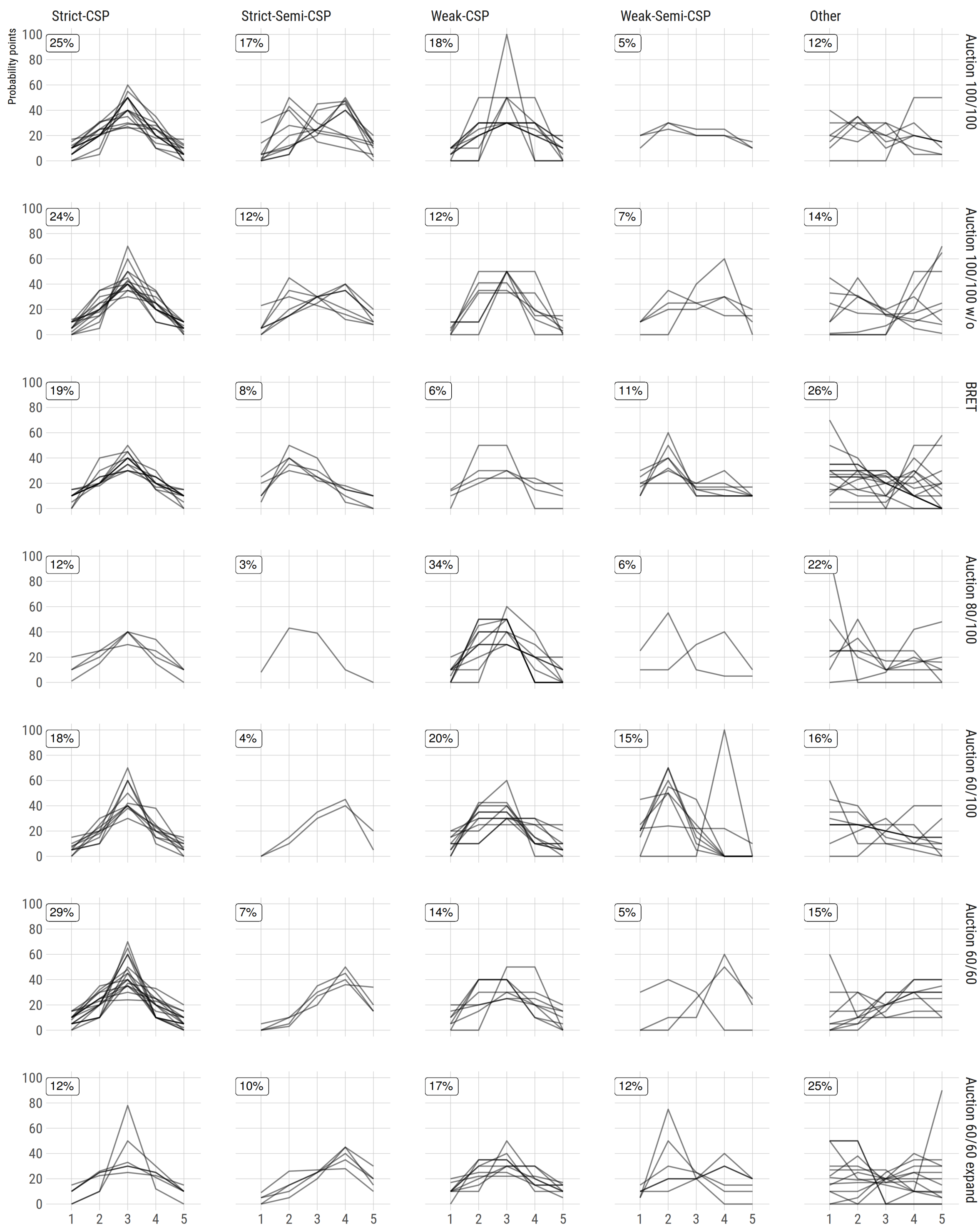

Figure 2: Non-uniform responses at a glance. Each piece-wise line represents a subject. The percentages are frequencies of the types within the treatment.

Table 3: Frequencies of peaked responses

Treatment	Strict-CSP	Weak-CSP	Strict-Semi-CSP	Weak-Semi-CSP	Subjects
Auction 100/100	15 (25.0%)	26 (43.3%)	25 (41.7%)	39 (65.0%)	60
Auction 100/100 w/o	14 (23.7%)	21 (35.6%)	21 (35.6%)	32 (54.2%)	59
BRET	12 (19.4%)	16 (25.8%)	17 (27.4%)	28 (45.2%)	62
Auction 80/100	4 (12.5%)	15 (46.9%)	5 (15.6%)	18 (56.3%)	32
Auction 60/100	10 (18.2%)	21 (38.2%)	12 (21.8%)	31 (56.4%)	55
Auction 60/60	17 (28.8%)	25 (42.4%)	21 (35.6%)	32 (54.2%)	59
Auction 60/60 expand	6 (11.5%)	15 (28.8%)	11 (21.2%)	26 (50.0%)	52
Pooled	78 (20.6%)	139 (36.7%)	112 (29.6%)	206 (54.4%)	379

Note: We list the number (and percentage) of responses that we categorize as having a strict CSP, a weak CSP, a strict semi-CSP, and a weak semi-CSP.

Table 4: Paired t-tests and signed rank tests of differences in weights between adjacent bins

	Bin 2 - Bin 1	Bin 3 - Bin 2	Bin 4 - Bin 3	Bin 5 - Bin 4
t-statistic	10.83	4.81	-6.68	-11.00
Corrected p-value	< 0.001	< 0.001	< 0.001	< 0.001
Signed rank z-statistic	11.01	5.14	-6.99	-11.18
Corrected p-value	< 0.001	< 0.001	< 0.001	< 0.001

Notes: We list t-statistic for a paired t-test and the signed rank test statistic for adjacent bins. We report the Bonferroni-corrected p-values of these tests. Since we have 4 pairwise tests, we multiply each p-value by 4. Each test is based on 379 observations.

or 5 is negative. The implication is that the weights tend to be increasing below Bin 3 and decreasing above Bin 3. In other words, this is evidence of a central single peak. Moreover, even with the Bonferroni correction, each test is significant at 0.001 level.

4 What drives the non-uniform responses?

In this section, we try to shed light on what drives the non-uniform responses, and particularly the CSP responses. In subsection 4.1, we show that these responses are not due to a possible misunderstanding of the objective distribution. In subsections 4.2 and 4.3, we examine the possibility that the non-uniform responses are an artefact of the experimental design in that, even if beliefs are uniform, the subject considers it preferable to give a non-uniform response. In 4.2, we ask whether the non-uniform responses might be driven by the way we implement the belief elicitation procedure. In 4.3, we analogously examine a potential role of payoff hedging. In subsection 4.4, we examine the extent to which the responses might reflect truly held beliefs, as opposed to noise. To do so, we analyze whether the responses are consistent either with bids or with the reported probabilities of winning the auction.

4.1 Possible problems with understanding the uniform distribution

In this subsection, we consider the possibility that subjects provide CSP responses because they do not understand that the true underlying distribution is uniform.

To begin with, note that after instructions are read aloud and displayed on each subject's screen, subjects have to clear a control question screen to be allowed to continue with the experiment (see section C.2 in Appendix C for the text of the control questions). Question 3 is devoted specifically to understanding the uniform distribution. It asks: *"What is the probability of your opponent's bid being in the range: a. 39 through 72; b. 22 though 47; c. 1 through 10; d. 16 through 56; e. 62 through 100"*. When subjects make a mistake in one or more questions, a pop-up screen tells them which questions they got wrong, and prompts them to change their answers. This procedure is iterated until no mistakes are submitted. We would therefore expect that a vast majority of subjects understand the environment, including the uniform distribution, when they start the experiment.

That being said, there is still some chance that subjects ultimately answer the control questions correctly by using some heuristics or trial and error, without fully understanding the environment. Such subjects are likely to provide incorrect *initial* responses to at least some control questions. On the other hand, those who provide correct initial responses to control questions might be considered to understand the relevant parts of the environment. Hence to examine the hypothesis that subjects provide CSP responses because they do not understand that the true underlying uniform distribution, we turn to evidence from control questions. We have data on the number of attempts, the number of mistakes on any question and sub-question, and the time spent on the control question screen. The time variable is an ambiguous indicator, but the number of mistakes constitutes a reliable proxy for *initial* understanding, and a noisy proxy for *ultimate* understanding. We particularly focus on question 3 that deals with understanding of the uniform distribution.

Subjects can fail one or more of the five sub-questions, and we keep track of each of them separately. The question generates a considerable number of initial mistakes: 58.8% subjects made at least one mistake in point a, 62% in b, 28% in c, 61.7% in d and 60.2% in e. The difference in the frequencies suggests that mistakes are to

Table 5: Relative frequency of responses by mistake type and χ^2 tests

	mistakes		wrong_uniform		wrong_simple_uniform	
	None	Some	None	Some	None	Some
N	59	320	104	275	272	107
Uniform	0.271	0.275	0.288	0.269	0.298	0.215
Weak-Semi-CSP	0.571	0.537	0.529	0.549	0.518	0.607
Extreme	0.114	0.129	0.115	0.131	0.129	0.121
Other	0.043	0.058	0.067	0.051	0.055	0.056
χ^2 test p-value	0.93		0.88		0.38	

some extent due to a problem with correctly taking into account the *boundaries* of the intervals. 72% of subjects correctly report the probability mass for the interval 1 through 10 as 10%, as immediately intuitive. However, when estimating the probability mass for the interval 16 through 56, a subtraction heuristic is largely used, making many subjects report 40% rather than the correct 41%.

We do not record the content of the wrong answers, only the number of wrong attempts. We use two different proxies of low comprehension. First, we identify subjects who make at least one mistake in any of the five sub-questions (`wrong_uniform = 1`). Second, in order to take into account the possibility that mistakes reflect the boundary problem rather than a basic misunderstanding of the uniform distribution, we identify those who make at least one mistake in the 1-through-10 sub-question (`wrong_simple_uniform = 1`). This sub-question is the easiest to answer and is immune to the boundary problem. For reference, we also identify subjects who make at least one mistake on any of the control questions on their first attempt (`mistakes = 1`).

Table 5 reports the relative frequencies of different types of responses for each of the two values of our three initial comprehension indicators. We use four mutually exclusive categories of responses: uniform, weak-semi-CSP (that includes as special cases all the centrally-concentrated response types), extreme (a weakly decreasing or a weakly increasing non-uniform response), and a residual "other" category. Response distributions are not significantly different for subjects committing zero rather than a positive number of overall mistakes, mistakes on any of the uniform distribution sub-questions, or specifically on the sub-question 3c.[21] Overall, we therefore conclude that the central bias identified in the responses is not due to a misunderstand-

[21] For subjects who initially fail the 1-through-10 sub-question, we see a slight decrease in the frequency of uniform responses and a corresponding increase in the frequency of CSP responses relative to subjects who do not initially fail this question. This suggests that the CSP reporting might after all be at least partly driven by a lack of understanding of the uniform distribution. We present additional analysis in Appendix B to evaluate robustness of the CSP reporting within each the two groups. The analysis shows that the central bias is robustly present in both groups. Hence even though misunderstanding of the uniform distribution, to the extent that it is present after clearing the control questions screen, might contribute to CSP reporting, it is not a major force behind such reporting.

ing of the uniform distribution.

4.2 Quadratic Scoring Rule and its implementation

In this subsection, we investigate the possibility that a subject holding the uniform subjective belief provides a non-uniform response because of incentives inherent in the QSR and its description. We make three points on the matter.

First, we do not provide subjects with the QSR payoff formula, but rather with advice on how to report (see subsection 2.2). Could it be that a subject holding the uniform subjective belief is somehow confused into thinking that a non-uniform response is superior? To evaluate this hypothesis, note that the main advice states that "*... in order to maximize your expected earnings from this procedure, you should report these likelihoods truthfully according to what you believe.*" Moreover, the second additional statement of advice (the only relevant one for someone with the uniform subjective belief) recommends reporting equal probabilities in bins that the subject believes to be equally likely. Therefore, we do not see why our description of the QSR would drive the deviations from the uniform response.

Second, it is well-known that the QSR gives risk-averse subjects an incentive to report a distribution that is "closer to" the uniform distribution than their true subjective beliefs (see, for example, Harrison et al. 2017). As a result, someone with both knowledge of the QSR payoff formula and understanding of its incentive properties might give a biased report of her subjective non-uniform belief. However, when the subjective belief is uniform, there is no bias in reporting. We state this result formally in Proposition 1 in Appendix A. Moreover, in comparison to Harrison et al. (2017), we prove that this result is true even for preferences with reasonable degrees of risk loving. Therefore, with the exception of degrees of risk loving that are not commonly observed, the non-uniform responses cannot be a consequence of deviations from risk neutrality if the underlying subjective belief is uniform.

Third, turning to a more behavioral perspective, given that we use an odd number of bins, the central bin might be salient or focal for putting a relatively large probability weight into. While we cannot rule out this possibility, we point out that Fairley, Parelman, Jones, and McKell (2019), who use 20 bins in a setting fundamentally similar to ours, report an auxiliary result that, on average, beliefs are single-peaked with the peak in bin 10.[22] Since 20 bins arguably provide a finer grid for reporting beliefs about a random variable, this finding suggests that the central single peak is not purely a consequence of a small, odd number of bins

[22] See their Figure 4 and A1. The authors do not point out the CSP property of the average beliefs.

in our experiment.

4.3 Hedging

Next, we investigate the possibility that subjects who hold true uniform subjective belief provide a non-uniform response because they use their payoff from the belief elicitation as a hedge for their payoff from the auction. Obviously, only a risk-averse subject desires to hedge.[23]

The first comment is analogous to that given in the previous subsection. Given that the instructions do not provide a payoff formula for the QSR and given the advice provided, the possibility of hedging is arguably not apparent to subjects. However, analogously the previous subsection, we consider how subjects who understand incentives for hedging would report given their subjective belief. In particular, we want to see whether the non-uniform responses could be due to risk-aversion-driven hedging despite the underlying subjective belief being uniform.

Intuitively, the auction does not pay off if and only if the bid of the computerized opponent is above the subject's bid. As a result, in order to hedge against losing the auction, a subject should bias the belief report toward the computerized bids in excess of her own bid and away from computerized bids below her own bid. We formalize this result in Proposition 2 in Appendix A. This proposition implies that if a risk-averter has the uniform subjective belief, then the reported probability in all bins below the bin in which her bid is located should be the same and strictly less than 20%, the reported probability in all bins above the bin in which her bid is located should be the same and strictly more than 20% and the reported probability for the bin in which her bid is located should be somewhere (weakly) in between. Among 379 subjects in our data, only 9 are strictly consistent with these predictions. Once we allow various probability reports below the bin or above the bin in which the bidder's bid is located to vary up to 1 percentage point[24] or 3 percentage points or 5 percentage points, this number increases to 10 or 11 or 13 (3.4% of the sample), respectively. Since this fraction is negligible compared to the fraction of subjects with non-uniform responses, we conclude that the non-uniform responses that we observe cannot be a consequence of payoff hedging if the underlying subjective belief is uniform.

[23]Recall that each subject is paid by the QSR for either eliciting beliefs of the opponent's bid or for eliciting beliefs of winning the auction, but not both. Therefore the payoff from the latter elicitation does not enter into consideration when thinking about hedging using the payoff from the former elicitation.

[24]This could occur, despite the subjects' best effort to equalize the reported probabilities, due to a perception that only integer probabilities can be reported (which is not the case).

4.4 Consistency of responses

In this subsection, we try to shed light on the question of whether the non-uniform responses, and particularly the CSP responses, capture truly-held subjective beliefs. If not, such responses could be noisy reports of underlying well-defined subjective beliefs, or they could reflect a lack of well-defined subjective beliefs or confusion, responses we refer to as "pure noise."

We begin by asking whether there is at least some evidence of the non-uniform responses being consistent with "external" behavior, namely bidding in our experiment. We define the response of a subject to be *externally consistent* if this subject's bid maximizes his expected utility (EU) under *some* (strictly increasing) utility function and under *some* belief about the opponent's bid that, when aggregated within each of the 5 bins, results in the given response.[25] Based on this definition, the only subjects who might have externally inconsistent responses are: (a) those whose bid exceeds the value when the probability of winning at this bid is strictly positive; (b) those whose bid results in a zero probability or a zero surplus if winning when there is a different bid at which the auction could be profitably won with a positive probability; or (c) those who bid below their value, but whose bid could be reduced without affecting a positive probability of winning. All subjects bid below their value, making (a) irrelevant. Only 3 subjects have a guaranteed zero probability of winning given their response and only 2 of these fit (b). Another 4 subjects fit (c). That is, only 6 out of 379 subjects (about 1.6% of the sample) display externally inconsistent responses. The remaining 373 subjects have externally consistent responses.

Given that the concept of external consistency imposes relatively little discipline on the responses, we also check for the relationship between responses and bids across subjects. This exercise is motivated by considering how the subject's EU-maximizing bid would change if the belief about the opponent's bid were to shift toward higher bids. Holding the utility function constant, such shift of beliefs might be expected to lead to a higher bid.[26] To examine the extent to which this theoretical comparative static is borne in between-subject

[25]This is, admittedly, a fairly non-restrictive definition. However, it is arguably the only possible one in the absence of knowledge of subject domain-specific risk attitude and in the absence of assumptions on how beliefs summarized by a response are distributed within the individual bins.

[26]This is not the case for all possible belief shifts to the right in the sense of first-order stochastic dominance. However, a simple sufficient, but not necessary, condition for the bid to (weakly) increase is that the after-shift belief cdf $G(\cdot)$ satisfies, relative to the pre-shift belief cdf $F(\cdot)$, that $G(b_1)/F(b_1) \leq G(b_2)/F(b_2)$ for any $b_1 < b_2$ such that $F(b_1) > 0$. To see this, let $u(\cdot)$ be the utility function normalized such that $u(0) = 0$. Then the expected payoff from bidding b when having the belief $F(\cdot)$ is $u(v-b)F(b)$. We then have that if for $b_1 < b_2$ with $F(b_1) > 0$ it is the case that $u(v-b_1)F(b_1) \leq u(v-b_2)F(b_2)$, it must also be the case that $u(v-b_1)G(b_1) \leq u(v-b_2)G(b_2)$. As a result, whenever b_2 generates at least as high an expected utility as b_1 under $F(\cdot)$, it does so also under $G(\cdot)$, and a similar implication analogously holds for strict inequalities as well. Hence the peak of the expected utility function cannot move to the left under G relative to F, and will typically move to the right.

comparisons, we rank-correlate across subjects their bid with the probability that the bid of the opponent does not exceed the upper boundary of Bin i implied by the response, repeating the exercise for $i \in \{1, .., 4\}$. To make this exercise meaningful as a reflection of the comparative static discussed above, we must assume that subjects reporting lower probabilities up to Bin i are not systematically more risk averse than subjects reporting higher probabilities. We perform the rank-correlation computations for four groups of subjects. In the first group, we use all 373 externally consistent subjects. In the second group, we use only the externally consistent subjects with non-uniform responses. Next, we split this group into subjects with Weak-Semi-CSP responses (the third group) and others (the fourth group). The resulting rank-correlations are presented in Table 6.

We observe that the rank correlations are negative and, with the exception of one case, statistically significant (even after a Bonferroni correction for multiple pairwise testing). That is, reporting lower probabilities up to Bin i (hence assigning higher probabilities above Bin i) is indeed significantly correlated with higher bids. This is a very broad and robust observation. It holds within all four considered groups, which comprise both wide and successively narrower sub-samples of the data. This observation provides further evidence that non-uniform responses in general, and CSP responses in particular, cannot be dismissed as pure noise. However, it is unclear to what extent the responses reflect true *ex ante* subjective beliefs that subjects hold before they decide on their bid. Given that subjects had to decide on their bids before being probed for their beliefs, their responses might also be driven by an *ex-post* justification of the submitted bids.

Since the concept of external consistency is not powerful in determining whether the responses capture truly-held subjective beliefs, or noisy reports of underlying well-defined subjective beliefs, or pure noise, we turn to a different concept of consistency. To streamline exposition, in the rest of this subsection we refer to one's belief report about the bid of the opponent as *response 1* (this is what we otherwise refer to as "response"), and we refer to one's winning probability report as *response 2*. We define the pair of responses of a subject to be *internally consistent* if there exists a belief about the opponent's bid such that: *a)* it can be aggregated within the 5 bins to reconstruct response 1; *b)* given these beliefs, the implied probability of winning the auction at the submitted bid is equal to response 2. If a subject's responses are internally consistent, there is an argument for considering such responses to capture truly held beliefs.

If a subject's responses are not internally consistent, we next ask whether they could be considered to be noisy reports of well-defined subjective beliefs. In such case, assuming that noise in the two responses is independent, we expect the two responses to become internally consistent once noise is removed from

Table 6: Spearman's rank-correlations of (scaled) bid with belief of the bid b_c of the opponent

Subject group	Statistic	$\Pr(b_c \leq 20)$	$\Pr(b_c \leq 40)$	$\Pr(b_c \leq 60)$	$\Pr(b_c \leq 80)$	Subjects
Externally consistent	Spearman's ρ	-0.241	-0.344	-0.286	-0.139	373
	Corrected p-value	< 0.001	< 0.001	< 0.001	0.028	
Non-uniform	Spearman's ρ	-0.303	-0.419	-0.372	-0.212	269
	Corrected p-value	< 0.001	< 0.001	< 0.001	0.002	
Weak-semi-CSP	Spearman's ρ	-0.215	-0.333	-0.317	-0.154	205
	Corrected p-value	0.008	< 0.001	< 0.001	0.110	
Other non-uniform	Spearman's ρ	-0.552	-0.634	-0.551	-0.476	64
	Corrected p-value	< 0.001	< 0.001	< 0.001	< 0.001	

Notes: We list the Spearman's ρ rank-correlations of (scaled) bid with belief that the bid b_c of the opponent does not exceed 20, 40, 60 or 80, respectively, for groups of subjects listed in the table. All listed groups exclude the 6 externally inconsistent subjects. In *auction 60/60* and *auction 60/60 expand*, we rescale both bids and belief bin boundaries by a factor of 100/60. We report the Bonferroni corrected p-values of these tests. Since we have 4 pairwise correlations for each group, we multiply each p-value by 4.

response 1 (*correction 1*) or response 2 (*correction 2*). If the pair of responses becomes internally consistent after at least one of the two corrections, we consider that subject to have well-defined subjective beliefs. In the opposite case, we consider that subject's beliefs to be pure noise.[27]

Applicability of this methodology requires to posit true subjective beliefs. If beliefs were allowed to be subject-specific, one could obtain a 100% internal consistency rate after correction 1 or after correction 2 by choosing a subjective belief that perfectly matches response 2 or response 1, respectively. Hence to render the exercise meaningful, we restrict attention to the case of the subjective beliefs being uniform. This choice is motivated by the shape of the objective distribution. This means that, under correction 1, response 1 is replaced by the uniform response. Analogously, under correction 2, response 2 is replaced by the objective probability of winning the auction at the submitted bid given the uniform distribution of the opponent's bid. As a consequence, we might end up classifying the responses of some internally inconsistent subjects with non-uniform subjective beliefs as pure noise as opposed to noisy reports of well-defined subjective beliefs just because we do not consider the correct subjective beliefs. Also, we might end up with such erroneous pure-noise classification even if the true subjective beliefs are uniform but the noise in the two responses is so large that none of the two corrections establishes internal consistency. Hence the incidence of pure noise responses in our data should be taken as an upper bound on the true incidence of not-well-defined beliefs.

Table 7 reports results of the consistency and correction classification. We observe that, overall, about 65% of responses are internally consistent. For the uniform-response subjects, it is almost 77%, whereas for the non-uniform-response subjects, it is about 60%. For the weak-semi-CSP subjects in particular, 66% of responses are internally consistent. Performing one or the other correction establishes internal consistency for about another quarter of responses in each of these three groups (for the uniform response group, this happens by construction). That is, (at least) about a quarter of the non-uniform responses in general, and the single-peaked responses in particular, can be thought of as being noisy reports of a well-defined underlying (uniform or non-uniform) subjective belief. Finally, (at most) about 14% of the non-uniform responses in general and 10% of the single-peaked responses in particular can be thought of as pure noise.

This exercise implies that a majority of non-uniform responses might indeed reflect true subjective beliefs. Only for a minority (at most one seventh) of non-uniform responses there is little evidence of any other explanation but pure noise. And as discussed above, even this fraction should be taken as an upper bound on the

[27]Performing both noise corrections simultaneously by definition establishes internal consistency. Therefore, we do not consider such exercise.

Table 7: Internal consistency of beliefs

Subject Group	Consistent	Inconsistent					Subjects
		All	Consistent after			Inconsistent after	
			Correction 1	Correction 2	Correction 1 or 2	either correction	
All	246 (64.9%)	133 (35.1%)	53 (14.0%)	50 (13.2%)	96 (25.3%)	37 (9.8%)	379
Uniform	80 (76.9%)	24 (23.1%)	0 (0.0%)	24 (23.1%)	24 (23.1%)	0 (0.0%)	104
Non-uniform	166 (60.4%)	109 (39.6%)	53 (19.3%)	26 (9.5%)	72 (26.2%)	37 (13.5%)	275
Weak-semi-CSP	136 (66.0%)	70 (34.0%)	34 (16.5%)	23 (11.2%)	50 (24.3%)	20 (9.7%)	206
Other non-uniform	30 (43.5%)	39 (56.5%)	19 (27.5%)	3 (4.4%)	22 (31.9%)	17 (24.6%)	69

Notes: We list the number (and percentage) of belief reports that we categorize as being internally consistent and internally inconsistent. We further break down the latter group into belief reports that become consistent after *correction 1*, after *Correction 2*, after *correction 1* or *correction 2*, and those that do not become consistent after either of the two corrections.

true incidence of pure noise responses in the data.

5 Implications for biases in Bayesian updating

In this section, we discuss implications of the central tendency bias on well-known biases in Bayesian updating when there are more than 2 ordered states of the world. It is well-known that subjects' updating behavior often does not conform to predictions of the Bayes' rule.[28] Specific individual characteristics may trigger different failures in Bayesian updating in the same decision task, for instance due to a personal inclination to overweight prior over new information (Achtziger, Alós-Ferrer, Hügelschäfer, and Steinhauser, 2014). We argue that the central tendency bias can help to rationalize at least two well-documented biases, namely conservatism and overprecision, even for the same decision maker, depending on the type of new information received.

The first bias is *conservatism*, defined as the tendency of subjects to overweight prior information and therefore to insufficiently adjust their posteriors to new information (Phillips and Edwards, 1966).[29] The second bias, *overprecision* (or Bayesian overconfidence), describes a situation in which the variance of posterior beliefs is lower than that justified by the acquired information.[30]

Rather than a feature of the updating process, however, conservatism and overprecision could be observationally equivalent to proper Bayesian updating under *ex ante* beliefs that are centrally biased relative to the assumed underlying distribution. The two biases can even coexist for the same decision maker, and which of the two occurs depends on the type of new information received. In particular, the central tendency bias suggests an explanation for conservatism in settings where the new information is relatively "extreme", i.e., supporting a state (or states) close to a boundary of the state space. If subjects perceive the prior distribution to be less variable than what the experimenter attempts to induce, we should observe, in response to extreme new information, a subjective posterior that is closer to the induced prior than the objective Bayesian posterior is.[31] In contrast, when the new information is "central", i.e., supporting centrally located states, the subjective

[28] A sample of this large literature would include Kahneman and Tversky (1973), Bar-Hillel (1980), Grether (1980, 1992), Weber (1994), Gigerenzer and Hoffrage (1995), El-Gamal and Grether (1995), Zizzo, Stolarz-Fantino, Wen, and Fantino (2000), Sedlmeier and Gigerenzer (2001), Kahneman and Frederick (2002), Charness, Karni, and Levin (2007, 2010), Oechssler, Roider, and Schmitz (2009), Achtziger and Alós-Ferrer (2013), Hawkins et al. (2015), and Cassey, Hawkins, Donkin and Brown (2016).

[29] Also see Beach (1968), Marks and Clarkson (1972), De Swart, J. H. (1972), Griffin and Tversky (1992), Erev, Wallsten, and Budescu (1994), Oechssler, Roider, and Schmitz (2009), and Corner, Harris, and Hahn (2010).

[30] For instance, there is evidence that subjects facing a uniform *ex-ante* distribution of events tend to hold updated beliefs that, even though unbiased on average, are characterized by a lower variance than implied by Bayesian updating (Harrison and Swarthout, 2019).

[31] See He and Xiao (2017) for a related model. Also, Benjamin, Rabin, and Raymond (2016) and Kovach (2015) propose other

posterior will exhibit overprecision relative to the objective Bayesian posterior.

To illustrate how our results can imply behavior consistent with conservatism and overprecision, consider the following example. Suppose that there are 5 ordered states of the world, $\omega_i = i$ for every $i \in \{1, ..., 5\}$. Also suppose that there is a uniform objective prior on the states, $Pr(\omega_i) = 0.2$ for every $i \in \{1, ..., 5\}$. The subject receives one signal, $s_i \in \{1, ..., 5\}$, which matches the true state with probability 0.8, i.e., $Pr(s_i|\omega_j) = 0.8$ if $i = j$. The signal does not match the state with probability 0.2, and there is a uniform distribution on signal realization over the remaining states, i.e., $Pr(s_i|\omega_j) = 0.05$ if $i \neq j$. Therefore, upon observing an extreme signal, say s_5, a Bayesian subject will report a posterior of

$$Pr(\omega_5|s_5) = \frac{Pr(s_5|\omega_5)Pr(\omega_5)}{\sum_{i=1}^{5} Pr(s_5|\omega_i)Pr(\omega_i)} = 0.8.$$

Further, the expected posterior value of the state will be

$$EV = \sum_{i=1}^{5} \omega_i * Pr(\omega_i|s_5) = 4.5.$$

Suppose, however, that the subject has an *ex-ante* subjective belief that, rather than being uniform, exhibits the central tendency bias. For an illustration, take $Pr(\omega_1) = 0.15$, $Pr(\omega_2) = 0.2$, $Pr(\omega_3) = 0.3$, $Pr(\omega_4) = 0.2$, and $Pr(\omega_5) = 0.15$. Upon observing the extreme signal s_5, the subject will report a posterior of $Pr(\omega_5|s_5) = 0.738$ and an expected value $EV = 4.38$. Since both are insufficiently sensitive to the extreme signal, the experimenter will conclude that the subject exhibits behavior consistent with conservatism.

Now consider the same example again, but this time assuming that the signal received by the same subject is s_3. The Bayesian posterior belief starting from the uniform prior should be:

$$Pr(\omega_1|s_3) = 0.05; Pr(\omega_2|s_3) = 0.05; Pr(\omega_3|s_3) = 0.8; Pr(\omega_4|s_3) = 0.05; Pr(\omega_5|s_3) = 0.05.$$

When the agent holds prior beliefs that exhibit the central tendency bias as assumed above, the posterior beliefs are instead:

$$Pr(\omega_1|s_3) \simeq 0.027; Pr(\omega_2|s_3) \simeq 0.036; Pr(\omega_3|s_3) \simeq 0.873; Pr(\omega_4|s_3) \simeq 0.036; Pr(\omega_5|s_3) \simeq 0.027.$$

While unbiased on average ($EV = 3$), these posterior beliefs are characterized by overprecision in that the

models of conservatism.

distribution is more concentrated than the Bayesian posterior given the correct prior (the variance is equal to $\sigma^2 \simeq 0.29$ instead of $\sigma^2 = 0.50$.)

Hence, the central tendency bias in prior beliefs, as found in our experiment, may offer an alternative explanation for purported failures in Bayesian updating in settings where there are more than 2 ordered states of the world. In other words, another explanation for conservatism and overprecision is that, rather than failing to process information according to the Bayes rule, subjects have subjective prior beliefs that are more concentrated in the middle of the ordered state space than the objective distribution.

Our experiment measures *ex-ante* beliefs only, and therefore is not equipped to test the implications of the central tendency bias on Bayesian updating. The nature of the exercise shown in this section is purely exploratory, but it is our view that it illustrates interesting implications and suggests a direction for promising future research.

6 Conclusion

In our experiment, we describe the distribution of bids of a random-bid auction opponent as being uniform. We subsequently elicit beliefs of this distribution using the Quadratic Scoring Rule. To our surprise, we find evidence that the reported beliefs are largely non-uniform and that many tend to be centrally biased.

The judgments literature has found such a central tendency bias in other settings. Our results suggest that the central tendency bias is more general than the collection of one-at-a-time judgments. The central tendency bias can also be observed when subjects report their beliefs of a probability distribution that is objectively uniform. Even more strikingly, this pattern occurs despite subjects demonstrating the ability to compute probabilities of events drawn from the uniform distribution. Moreover, many of the central tendency bias experiments are not incentivized, whereas we find evidence of this bias in an incentivized belief elicitation procedure.

We argue and empirically demonstrate that the central tendency bias found in our data is not an artefact of our experimental design. In contrast, there is some indicative evidence that the reported beliefs exhibiting the central tendency bias capture, or at least approximate, truly held beliefs for most of the subjects. This evidence draws upon consistency of reported beliefs with another belief report and with the auction bid. If confirmed by future research, such finding would imply that it is not easy to control beliefs in the laboratory because, in line with the insights of the decision-from-experience literature (see for instance Cohen, Plonsky and Erev 2020), subjects having read and understood the instructions can still hold idiosyncratic subjective beliefs.

Our data does not allow us to shed light on the causes of the documented central tendency bias. We speculate that there could be several potential explanations. For example, subjects might over-generalize from naturally occurring situations. Uniform distributions, while widespread in man-made settings like casino games and lotteries, are rare in the natural world, where central singe-peaked distributions are more common. Or, alternatively, subjects might find it beneficial to be 'in the center' since they partially confuse the distribution with its mean or median.[32]

The central tendency bias offers an explanation for well-known biases in Bayesian updating, such as conservatism and overprecision. In studies of biases in Bayesian updating, it is usually assumed that subjects correctly internalize the induced prior distribution. Our results suggest that this is not necessarily the case. Conservatism and overprecision can also be rationalized by subjects holding prior beliefs that are more concentrated in the middle of the ordered state space than the objective distribution. The central tendency bias also suggests that probabilistic forecasts (rather than point forecasts) of events might have less variance than the actual events. While there is mixed evidence of this in macroeconomic forecasts (Smyth and Ash, 1981; Stekler, 1975),[33] we are interested to learn whether this implication of our results can be found in studies of predictions of uncertain events.

Although we observe the central tendency bias in seven different experimental settings, we are interested to learn the extent to which our results are robust to different stochastic distributions, different elicitation specifications (for instance, different numbers of bins or bins that do not have identical sizes), and other experimental details. We hope that future experimental work can shed light on the extent to which our results are robust.

[32] An observationally equivalent outcome follows from targeting the distribution of the sample mean. Since the objective distribution is uniform, the distribution of the sample mean based on 2 or more draws in indeed single-peaked. Nonetheless, in our setting subjects perform a single draw and do so *after* the belief elicitation phase, so this point should not apply.

[33] Engelberg, Manski, and Williams (2009) also examine probabilistic forecasts of experts but do not compare the variances of the predictions with the actually realized variances.

References

Achtziger, Anja, and Alós-Ferrer, Carlos (2013): ”Fast or rational? A response-times study of Bayesian updating,” *Management Science*, 60(4), 923–938.

Achtziger, Anja, Alós-Ferrer, Carlos, Hügelschäfer, Sabine, and Steinhauser, Marco (2014): ”The neural basis of belief updating and rational decision making,” *Social Cognitive and Affective Neuroscience*, 9(1), 55–62.

Akaike, Hirotugu (1974): ”A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Algom, Daniel and Marks, Lawrence E. (1990): ”Range and regression, loudness scales, and loudness processing: Toward a context-bound psychophysics,” *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 706–727.

Allen, Franklin (1987): ”Discovering personal probabilities when utility functions are unknown,” *Management Science*, 33(4), 542–544.

Andersen, Steffen, Fountain, John, Harrison, Glenn W., and Rutström, E. Elisabet (2014): ”Estimating subjective probabilities,” *Journal of Risk and Uncertainty*, 48(3), 207–229.

Arad, Ayala and Penczynski, Stefan (2018): ”Multi–Dimensional Reasoning in Competitive Resource Allocation Games: Evidence from Intra–Team Communication,” Working Paper Tel Aviv University.

Arad, Ayala and Rubinstein, Ariel (2012): ”Multi-dimensional iterative reasoning in action: The case of the Colonel Blotto game,” *Journal of Economic Behavior and Organization*, 84(2), 571–585.

Armantier, Olivier and Treich, Nicolas (2013): ”Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging,” *European Economic Review*, 62, 17–40.

Bar-Hillel, Maya (1980): ”The base-rate fallacy in probability judgments,” *Acta Psychologica*, 44(3), 211–233.

Beach, Lee (1968): ”Probability magnitudes and conservative revision of subjective probabilities,” *Journal of Experimental Psychology*, 77(1), 57–63.

Benjamin, Daniel J., Rabin, Matthew, and Raymond, Collin (2016): "A model of nonbelief in the law of large numbers," *Journal of the European Economic Association*, 14(2), 515–544.

Blanco, Mariana, Engelmann, Dirk, Koch, Alexander K., and Normann, Hans-Theo (2010): "Belief elicitation in experiments: Is there a hedging problem?" *Experimental Economics*, 13(4), 412–438.

Brier, Glenn W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 78(1), 1–3.

Camerer, Colin and Weigelt, Keith (1988): "Experimental tests of a sequential equilibrium reputation model," *Econometrica*, 56(1), 1–36.

Cassey, Peter, Hawkins, Guy E., Donkin, Chris, and Brown, Scott D. (2016): "Using alien coins to test whether simple inference is Bayesian," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(3), 497–503.

Charness, Gary, Karni, Edi, and Levin, Dan (2007): "Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance," *Journal of Risk and Uncertainty*, 35(2), 129–148.

Charness, Gary, Karni, Edi, and Levin, Dan (2010): "On the conjunction fallacy in probability judgment: New experimental evidence regarding Linda," *Games and Economic Behavior*, 68(2), 551–556.

Chernev, Alexander (2004): "Extremeness aversion and attribute-balance effects in choice," *Journal of Consumer Research*, 31(2), 249–263.

Cohen, Doron, Plonsky, Ori, and Erev, Ido (2020): "On the impact of experience on probability weighting in decisions under risk," *Decision*, Advance online publication," https://doi.org/10.1037/dec0000118

Corner, Adam, Harris, Adam, and Hahn, Ulrike (2010): "Conservatism in belief revision and participant skepticism," *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32, 1625–1630.

Crawford, L. Elizabeth (2019): "Reply to Duffy and Smith's (2018) reexamination," *Psychonomic Bulletin and Review*, 26(2), 693–698.

Crosetto, Paolo and Filippin, Antonio (2013): ”The ’bomb’ risk elicitation task,” *Journal of Risk and Uncertainty*, 47(1), 31–65.

De Swart, J. H. (1972): ”Effects of diagnosticity and prior odds on conservatism in a bookbag-and-pokerchip situation,” *Acta Psychologica*, 36(1), 16–31.

Duffy, Sean, Gussman, Steven, and Smith, John (2019): ”Judgments of length in the economics laboratory: Are there brains in choice?” Working paper, Rutgers University-Camden.

Duffy, Sean, Huttenlocher, Janellen, Hedges, Larry V., and Crawford, L. Elizabeth (2010): ”Category effects on stimulus estimation: Shifting and skewed frequency distributions,” *Psychonomic Bulletin and Review*, 17(2), 224–230.

Duffy, Sean and Smith, John (2018): ”Category effects on stimulus estimation: Shifting and skewed frequency distributions-A reexamination,” *Psychonomic Bulletin and Review*, 25(5), 1740–1750.

Duffy, Sean and Smith, John (2019): ”On the Category Adjustment Model: Another look at Huttenlocher, Hedges, and Vevea (2000),” Working paper, Rutgers University-Camden.

Duffy, Sean and Smith, John (2020): ”Omitted-variable bias and other matters in the defense of the category adjustment model: A reply to Crawford (2019),” *Journal of Behavioral and Experimental Economics*, forthcoming.

El-Gamal, Mahmoud A. and Grether, David M. (1995): ”Are people Bayesian? Uncovering behavioral strategies,” *Journal of the American Statistical Association*, 90(432), 1137–1145.

Engelberg, Joseph, Manski, Charles F., and Williams, Jared (2009): ”Comparing the point predictions and subjective probability distributions of professional forecasters,” *Journal of Business and Economic Statistics*, 27(1), 30–41.

Erev, Ido, Wallsten, Thomas S., and Budescu, David V. (1994): ”Simultaneous over-and underconfidence: The role of error in judgment processes,” *Psychological Review*, 101(3), 519–527.

Fairley, Kim, Parelman, Jacob M., Jones, Matt, and McKell, Carter, R. (2019): ”Risky health choices and the Balloon Economic Risk Protocol,” *Journal of Economic Psychology*, 73, 15–33.

Gigerenzer, Gerd and Hoffrage, Ulrich (1995): ”How to improve Bayesian reasoning without instruction: Frequency formats,” *Psychological Review*, 102(4), 684–704.

Grether, David M. (1980): ”Bayes rule as a descriptive model: The representativeness heuristic,” *Quarterly Journal of Economics*, 95(3), 537–557.

Grether, David M. (1992): ”Testing Bayes rule and the representativeness heuristic: Some experimental evidence,” *Journal of Economic Behavior and Organization*, 17(1), 31–57.

Griffin, Dale and Tversky, Amos (1992): ”The weighing of evidence and the determinants of confidence,” *Cognitive Psychology*, 24(3), 411–435.

Harrison, Glenn W., Martínez-Correa, Jimmy, Swarthout, J. Todd, and Ulm, Eric R. (2017): ”Scoring rules for subjective probability distributions,” *Journal of Economic Behavior and Organization*, 134, 430–448.

Harrison, Glenn W. and Swarthout, J. Todd (2019): ”Belief Distributions, Overconfidence and Bayes Rule,” Working paper, *Georgia State University*.

Hassidim, Avinatan, Marciano, Deborah, Romm, Assaf, and Shorrer, Ran I. (2017): ”The Mechanism Is Truthful, Why Aren't You?” *American Economic Review*, 107(5), 220–224.

Hawkins, Guy E., Hayes, Brett K., Donkin, Chris, Pasqualino, Martina, and Newell, Ben R. (2015): ”A Bayesian latent-mixture model analysis shows that informative samples reduce base-rate neglect,” *Decision*, 2(4), 306–318.

He, Xue Dong and Xiao, Di (2017): ”Processing consistency in non-Bayesian inference,” *Journal of Mathematical Economics*, 70, 90–104.

Hollingworth, Harry Levi (1910): ”The central tendency of judgment,” *Journal of Philosophy, Psychology and Scientific Methods*, 7(17), 461–469.

Hossain, Tanjim and Okui, Ryo (2013): ”The binarized scoring rule,” *Review of Economic Studies*, 80(3), 984–1001.

Huttenlocher, Janellen, Hedges, Larry V., and Duncan, Susan (1991): ”Categories and particulars: Prototype effects in estimating spatial location,” *Psychological Review*, 98(3), 352–376.

Huttenlocher, Janellen, Hedges, Larry V., and Vevea, Jack L. (2000): ”Why do categories affect stimulus judgment?” *Journal of Experimental Psychology: General*, 129, 220–241.

Jazayeri, Mehrdad and Shadlen, Michael N. (2010): ”Temporal context calibrates interval timing,” *Nature Neuroscience*, 13(8), 1020–1026.

Jones, Lynette A. and Hunter, Ian W. (1982): ”Force sensation in isometric contractions: a relative force effect,” *Brain Research*, 244(1), 186–189.

Jou, Jerwen, Leka, Gary E., Rogers, Dawn M., and Matus, Yolanda E. (2004): ”Contraction bias in memorial quantifying judgment: Does it come from a stable compressed memory representation or a dynamic adaptation process?” *American Journal of Psychology*, 117 (4), 543–564.

Kahneman, Daniel and Frederick, Shane (2002): ”Representativeness revisited: Attribute substitution in intuitive judgment,” In Gilovich, Thomas, Griffin, Dale, and Kahneman, Daniel (Eds.), *Heuristics and biases: The Psychology of Intuitive Judgment*, Cambridge University Press, 49–81.

Kahneman, Daniel and Tversky, Amos (1973): ”On the psychology of prediction,” *Psychological Review*, 80(4), 237–251.

Kahneman, Daniel and Tversky, Amos (1979): ”Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47 (2), 263–292.

Karni, Edi (2009): ”A mechanism for eliciting probabilities,” *Econometrica*, 77(2), 603–606.

Kareev, Yaakov, Arnon, Sharon, and Horwitz-Zeliger, Reut (2002): ”On the misperception of variability,” *Journal of Experimental Psychology: General*, 131(2), 287–297.

Kovach, Matthew (2015): ”Sticky Beliefs: A Characterization of Conservative Updating,” Working paper Virginia Tech.

Manski, Charles F. (2004): ”Measuring expectations,” *Econometrica*, 72(5), 1329–1376.

Marchiori, Davide, Di Guida, Sibilla, and Erev, Ido (2015): ”Noisy retrieval models of over-and undersensitivity to rare events,” *Decision*, 2(2), 82–106.

Marks, D. F. and Clarkson, J. K. (1972): "An explanation of conservatism in the bookbag-and-pokerchips situation," *Acta Psychologica*, 36(2), 145–160.

Matheson, James E. and Winkler, Robert L. (1976): "Scoring rules for continuous probability distributions," *Management Science*, 22(10), 1087–1096.

Neumann, Nico, Böckenholt, Ulf, and Sinha, Ashish (2016): "A meta-analysis of extremeness aversion," *Journal of Consumer Psychology*, 26(2), 193–212.

Oechssler, Jörg, Roider, Andreas, and Schmitz, Patrick W. (2009): "Cognitive abilities and behavioral biases," *Journal of Economic Behavior and Organization*, 72(1), 147–152.

Offerman, Theo, Sonnemans, Joep, Van de Kuilen, Gijs, and Wakker, Peter P. (2009): "A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes," *Review of Economic Studies*, 76(4), 1461–1489.

Phillips, Lawrence D. and Edwards, Ward (1966): "Conservatism in a simple probability inference task," *Journal of Experimental Psychology*, 72(3), 346–354.

Poulton, E. Christopher (1979): "Models for biases in judging sensory magnitude," *Psychological Bulletin*, 86(4), 777–803.

Qiu, Jianying and Weitzel, Utz (2016): "Experimental evidence on valuation with multiple priors," *Journal of Risk and Uncertainty*, 53(1), 55–74.

Radvansky, Gabriel A., Carlson-Radvansky, Laura A., and Irwin, David E. (1995): "Uncertainty in estimating distances from memory," *Memory and Cognition*, 23(5), 596–606.

Rubinstein, Ariel, Tversky, Amos, and Heller, Dana (1997): "Naive strategies in competitive games," In Albers, Wulf, Güth, Werner, Hammerstein, Peter, Moldovanu, Benny, van Damme, Eric (Eds.), *Understanding strategic interaction*, Springer, Berlin, Heidelberg, 394–402.

Savage, Leonard J. (1971): "Elicitation of personal probabilities and expectations," *Journal of the American Statistical Association*, 66(336), 783–801.

Schlag, Karl H., Tremewan, James, and Van der Weele, Joël J. (2015): "A penny for your thoughts: a survey of methods for eliciting beliefs," *Experimental Economics*, 18(3), 457–490.

Schlag, Karl H. and van der Weele, Joël (2013): ”Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality,” *Theoretical Economics Letters*, 3, 38–42.

Schlag, Karl H. and van der Weele, Joël J. (2015): ”A method to elicit beliefs as most likely intervals,” *Judgment and Decision Making*, 10(5), 456–468.

Schotter, Andrew and Trevino, Isabel (2014): ”Belief elicitation in the laboratory,” *Annual Review of Economics*, 6(1), 103–128.

Sedlmeier, Peter and Gigerenzer, Gerd (2001): ”Teaching Bayesian reasoning in less than two hours,” *Journal of Experimental Psychology: General*, 130(3), 380–400.

Simonson, Itamar and Tversky, Amos (1992): ”Choice in context: Tradeoff contrast and extremeness aversion,” *Journal of Marketing Research*, 29(3), 281–295.

Smyth, David J. and Ash, J. C. K. (1981): ”The underestimation of forecasts and the variability of predictions and outcomes,” *Bulletin of Economic Research*, 33(1), 37–44.

Stekler, Herman O. (1975): ”Why do forecasters underestimate?” *Economic Inquiry*, 13(3), 445–449.

Stevens, S. S. and Greenbaum, Hilda B. (1966): ”Regression effect in psychophysical judgment,” *Perception and Psychophysics*, 1(5), 439–446.

Weber, Elke U. (1994): ”From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events,” *Psychological Bulletin*, 115(2), 228–242.

Winkler, Robert L. and Murphy, Allan H. (1970): ”Nonlinear utility and the probability score,” *Journal of Applied Meteorology*, 9(1), 143–148.

Zizzo, Daniel John, Stolarz-Fantino, Stephanie, Wen, Julie, and Fantino, Edmund (2000): ”A violation of the monotonicity axiom: Experimental evidence on the conjunction fallacy,” *Journal of Economic Behavior and Organization*, 41(3), 263–276.

Appendices

A Propositions

A.1 Incentive compatibility of the QSR and risk non-neutrality

Suppose that the state space is partitioned into n mutually exclusive and exhaustive events (5 in our setting). Let $p_1, ..., p_n$ be the subjective probabilities and let $r_1, ..., r_n$ be the reported probabilities of the individual events. Any report must satisfy $r_1 + ... + r_n = 1$. The realized payoff is given by a constant minus a penalty linear in the square of the Euclidian distance between the realized and the reported probability vector. The former one is given by $(0, ..0, 1, 0, .., 0)$, with "1" on the position of the realized state. That is, the payoff if state i is realized is given by

$$\begin{aligned}\pi_i &\equiv \alpha - \beta \left((1 - r_i)^2 + \sum_{j \neq i} (0 - r_j)^2 \right) \\ &= (\alpha - \beta) + 2\beta r_i - \beta \sum_{j=1}^{n} r_j^2\end{aligned}$$

with $\alpha, \beta > 0$. Note that the minimum possible value of π_i is $\alpha - 2\beta$ and it is attained if $r_j = 1$ for some $j \neq i$, whereas the maximum possible value of π_i is α and it is attained if $r_i = 1$. In our setting $\alpha = 20$ and $\beta = 10$, implying a possible payoff range from 0 to 20 ECUs. Consider an EU maximizer with a strictly increasing and twice continuously differentiable utility function $u(\cdot)$. We then have the following result:[34]

Proposition 1 *Suppose that $u'(\cdot)$ is bounded away from 0 and ∞ on $[\alpha - 2\beta, \alpha]$. Then any optimal report $(r_1^*, .., r_n^*)$ satisfies:*

1. *for any event i, $r_i^* = 0$ if and only if $p_i = 0$, $0 < r_i^* < 1$ if and only if $0 < p_i < 1$ and $r_i^* = 1$ if and only if $p_i = 1$;*
2. *for any two distinct events i and j, if $p_i > p_j$, then $r_i^* > r_j^*$; by a contrapositive, if $r_i^* = r_j^*$, then $p_i = p_j$;*

[34] This proposition is similar to the theoretical results presented in Harrison et al. (2017). However, there are three novelties here: (1) the first statement in part 2; while the same result follows from Lemma 1 in Harrison et al. (2017) under risk aversion, we prove it generally for any risk preference (the contrapositive statement in part 2 is identical to Lemma 4 in Harrison et al., 2017); (2) the upper bound on risk-loving in part 3; while Harrison et al. (2017) prove the same result under risk aversion (Lemma 3), we partially extend it to risk loving as well; (3) part 4; in comparison to Proposition 4 in Harrison et al. (2017), this is a different way of expressing the result that risk aversion leads to a "flattening" of reported beliefs toward equal reports, whereas risk loving has the opposite effect; implications of the two approaches for data inference are arguably identical, though.

3. *for any two distinct events i and j, if $u''(x)/u'(x) \leq \beta^{-1}$ for any $x \in [\alpha - 2\beta, \alpha]$ and $p_i = p_j$, then $r_i^* = r_j^*$;*

4. *for any two distinct events i and j with $p_i > p_j > 0$, if $u(\cdot)$ is linear, then $r_i^* = p_i$ and $r_j^* = p_j$; if $u(\cdot)$ is strictly concave, then $r_i^*/r_j^* < p_i/p_j$; and if $u(\cdot)$ is strictly convex,*

   then $r_i^*/r_j^* > p_i/p_j$.

The most important implication of this proposition is that if a subject believes that all 5 bins are equally likely, then, barring a sufficiently high risk-loving, it is optimal to provide the uniform response.[35] That is, the non-uniform responses that we observe cannot be a consequence of deviations from risk neutrality if the underlying subjective belief is uniform.

Proof of Proposition 1. The decision-maker solves

$$\max_{r_1,..,r_n \in [0,1]} \sum_{i=1}^{n} p_i u \left(\alpha - \beta + 2\beta r_i - \beta \sum_{j=1}^{n} r_j^2 \right) \text{ s.t. } r_1 + ... + r_n = 1.$$

At the beginning, we are going to ignore the constraint $r_1 + ... + r_n = 1$ and

focus on the resulting "unconstrained" problem subject only to the usual

probability bounds of 0 and 1. Note that π_i is strictly increasing

and concave in r_i on $[0, 1]$ with $\partial \pi_i / \partial r_i |_{r_i=1} = 0$, whereas π_j, $j \neq i$, is strictly decreasing and concave in r_i on $[0, 1]$ with $\partial \pi_j / \partial r_i |_{r_i=0} = 0$. This implies that,

starting from $r_i = 0$, a small increase in r_i has a positive first-order

effect on π_i without any counterbalancing negative first order effect

on π_j, $j \neq i$. Since $u'(\cdot)$ is bounded away from 0

and ∞, this implies that $r_i^* > 0$ if $p_i > 0$. Likewise, starting

from $r_i = 1$, a small decrease in r_i has a positive first-order effect on π_j, $j \neq i$, without any counterbalancing first order effect on π_i, implying that $r_i^* < 1$ if $p_i < 1$. Moreover, if $p_i = 0$, it is trivial to see that $r_i^* = 0$ even without considering the lower probability bound of 0, and if $p_i = 1$, it is trivial to see that $r_i^* = 1$ even without considering the upper probability bound of 1. This series of observations jointly implies the equivalent of part

[35]To gauge what is meant by "sufficiently high risk-loving", note that by part 3 of Proposition 1, the result of equal reporting of equal true probabilities goes through for any utility function with constant absolute risk loving of at most $\beta^{-1} = 0.1$ if payoffs are denominated in ECUs, or of at most 0.5 if payoffs are denominated in Euros. Working with the constant absolute risk aversion utility function $u(x) = e^{0.1x}$ when counting in ECUs ($u(x) = e^{0.5x}$ when counting in Euros), and considering a 50/50 lottery between the two most extreme possible payoffs from the payoff elicitation, namely 0 and 20 (€0 and €4), the certainty equivalent of this lottery for a decision maker with such utility function is approximately 14.34 ECUs (€2.87). Arguably, few subjects are as risk-loving as this.

1 of the Proposition for the unconstrained problem. It also implies that the probability bounds of 0 and 1 are never binding in this problem and can therefore be ignored.

For the equivalent of part 2 of the Proposition for the unconstrained problem, suppose that $p_i > p_j$ and, by contradiction, $r_i^* \leq r_j^*$.

First, suppose that $r_i^* = r_j^*$. By part 1 of the Proposition, it then

must be the case that $0 < r_i^* = r_j^* < 1$. Now, starting from this point, consider a small increase in r_i and an exactly offsetting small decrease

in r_j. This change has a positive first-order effect on π_i, an offsetting negative first-order effect on π_j of the same absolute size

and no first-order effect on π_k for $k \neq i, j$. Since $p_i > p_j$ and $u'(\cdot)$ is bounded away from 0 and ∞, such perturbation increases EU, contradicting optimality of r^*. Second, suppose that $r_i^* < r_j^*$, implying that $\pi_i^* < \pi_j^*$. Now consider resetting r_i and r_j such that $r_i = r_j^*$ and $r_j = r_i^*$. This change does not affect π_k for $k \neq i, j$. As a result, the EU changes by

$$[p_i u(\pi_j^*) + p_j u(\pi_i^*)] - [p_i u(\pi_i^*) + p_j u(\pi_j^*)] = (p_i - p_j)[u(\pi_j^*) - u(\pi_i^*)] > 0.$$

But this means that such a change increases EU, contradicting optimality of r^*. Hence if $p_i > p_j$, it must be the case that $r_i^* > r_j^*$.

Since the probability bound constraints can be ignored as argued above, any

solution $(r_1^*, .., r_n^*)$ to the unconstrained problem must satisfy the

usual first-order necessary conditions

$$2\beta p_k u'(\pi_k^*) - 2\beta r_k^* \sum_{j=1}^{n} p_j u'(\pi_j^*) = 0, \quad k = 1, .., n, \tag{1}$$

where π_i^* is the resulting payoff if state i is realized. Summing

these n conditions gives

$$2\beta \sum_{k=1}^{n} p_k u'(\pi_k^*) - 2\beta \left(\sum_{k=1}^{n} r_k^* \right) \sum_{j=1}^{n} p_j u'(\pi_j^*) = 0,$$

implying that $r_1^* + ... + r_n^* = 1$. As a result, any solution to the unconstrained problem also solves the constrained problem. This observation hence proves the first two parts of the proposition also for the constrained problem.

For part 3 of the Proposition, if $p_i = p_j = 0$, then $r_i^* = r_j^* = 0$ is implied by part 1 of the Proposition. Now suppose that $p_i = p_j \in (0, 0.5]$.

Then part 1 of the Proposition implies that $r_i^*, r_j^* \in (0,1)$. By (1), it then must be the case that

$$\frac{r_i^*}{r_j^*} = \frac{u'(\pi_i^*)}{u'(\pi_j^*)}. \tag{2}$$

Clearly, if $r_i^* = r_j^*$, then $\pi_i^* = \pi_j^*$, and (2) is satisfied. Next, we are going to show that (2) cannot be satisfied for $r_i^* \neq r_j^*$. Suppose, by contradiction, that

there exists a report profile r^* with, without loss of generality, $r_i^* > r_j^*$ and $r_i^*, r_j^* \in (0,1)$ such that (2) is satisfied. Let $\bar{r} \equiv (r_i^* + r_j^*)/2$. Now consider a different report profile r' which differs from r^* by both r_i^* and r_j^* being replaced by $\bar{r}$, while the other reported probabilities are left unchanged. Now, starting from r', gradually increase r_i, by

the same amount gradually decrease r_j, and keep all the other reported probabilities fixed until r^* is reached. At any point along this trajectory, we have

that

$$\begin{aligned}
d\left(\ln \frac{r_i}{r_j}\right) &= d \ln r_i - d \ln r_j \\
&= \frac{dr_i}{r_i} + \frac{-dr_j}{r_j} \\
&= \frac{dr_i}{r_i} + \frac{dr_i}{r_j} \\
&= dr_i \left(\frac{1}{r_i} + \frac{1}{r_j}\right) \\
&\geq 4dr_i.
\end{aligned}$$

The last inequality follows from the constraint $r_i + r_j \leq 1$. Equality

potentially applies only at r', otherwise a strict inequality applies. At any point along this trajectory, we also have that

$$
\begin{aligned}
d\left[\ln \frac{u'(\pi_i)}{u'(\pi_j)}\right] &= d \ln u'(\pi_i) - d \ln u'(\pi_j) \\
&= \frac{u''(\pi_i)}{u'(\pi_i)} d\pi_i + \frac{u''(\pi_j)}{u'(\pi_j)}(-d\pi_j). \qquad (3)
\end{aligned}
$$

Now note that

$$
\begin{aligned}
d\pi_i &= 2\beta dr_i - 2\beta(r_i dr_i + r_j dr_j) \\
&= 2\beta dr_i - 2\beta(r_i - r_j) dr_i \\
&= 2\beta(1 - r_i + r_j) dr_i
\end{aligned}
$$

and

$$
\begin{aligned}
-d\pi_j &= 2\beta(-dr_j) + 2\beta(r_i dr_i + r_j dr_j) \\
&= 2\beta dr_i + 2\beta(r_i - r_j) dr_i \\
&= 2\beta(1 + r_i - r_j) dr_i,
\end{aligned}
$$

implying that $d\pi_i > 0$, $-d\pi_j > 0$ and

$$
d\pi_i - d\pi_j = 4\beta dr_i.
$$

These results and the upper bound of risk loving preferences assumed in part 3 of the Proposition imply that in (3) we have that

$$
\begin{aligned}
d\left[\ln \frac{u'(\pi_i)}{u'(\pi_j)}\right] &= \frac{u''(\pi_i)}{u'(\pi_i)} d\pi_i + \frac{u''(\pi_j)}{u'(\pi_j)}(-d\pi_j) \\
&\leq \beta^{-1}(d\pi_i - d\pi_j) \\
&= 4 dr_i.
\end{aligned}
$$

Overall, we therefore have at any point along the trajectory from r' to r^* that

$$
d\left[\ln \frac{r_i}{r_j} - \ln \frac{u'(\pi_i)}{u'(\pi_j)}\right] \geq 0,
$$

with equality potentially applying only at r' and a strict inequality applying otherwise. But this result and (2) then imply that

$$\frac{\bar{r}}{\bar{r}} < \frac{u'(\bar{\pi})}{u'(\bar{\pi})},$$

where

$$\bar{\pi} \equiv \alpha - \beta + 2\beta\bar{r} - \beta\left(\sum_{k\neq i,j} r_k^{*2} + 2\bar{r}^2\right),$$

a clear contradiction. Therefore if $p_i = p_j \in (0, 0.5]$, then, under the assumption of part 3, it must be the case that $r_i^* = r_j^*$.

For part 4 of the Proposition, if $p_i > p_j > 0$, it follows from parts 1 and 2 of the Proposition that $r_i^* > r_j^* > 0$. If $u(\cdot)$ is linear, then the result that $r_i^* = p_i$ and $r_j^* = p_j$ follows directly from (1) for events i and j. If $u(\cdot)$ is non-linear, then (1) implies that

$$\frac{r_i^*}{r_j^*} = \frac{p_i}{p_j}\frac{u'(\pi_i^*)}{u'(\pi_j^*)}.$$

Since $r_i^* > r_j^*$, it must be the case that that $\pi_i^* > \pi_j^*$. Hence if $u(\cdot)$ is strictly concave, $u'(\pi_i^*)/u'(\pi_j^*) < 1$, implying that $r_i^*/r_j^* < p_i/p_j$, whereas if $u(\cdot)$ is strictly convex, $u'(\pi_i^*)/u'(\pi_j^*) > 1$, implying that $r_i^*/r_j^* > p_i/p_j$. ■

A.2 Hedging

We follow the structure introduced in the previous subsection, but this time adapted to the parameterization used in the experiment. That is, $n = 5$, $\alpha = 20$ and $\beta = 10$. Moreover, the events now correspond to the individual bins, with event i corresponding to bin i, $i = 1, .., 5$. Let k be the bin that contains the bid. Also, let s be the auction surplus, that is, the difference between value and bid. We assume that this surplus is strictly positive.[36] For $i < k$, the belief elicitation payoff is augmented by the auction surplus, resulting in the overall payoff of

$$\pi_i \equiv s + 10 + 20r_i - 10\sum_{j=1}^{5} r_j^2.$$

An analogous payoff expression applies also to bin k if the bid of the opponent does not exceed the bid of the subject. Denote this payoff π_{kH} and let q be the subjective probability of such outcome conditional on the bid of the opponent being in bin k. For $i > k$, the belief elicitation payoff is *not* augmented by the auction surplus,

[36]This is the case for all 379 subjects in our sample.

resulting in the overall payoff of

$$\pi_i \equiv 10 + 20r_i - 10\sum_{j=1}^{5} r_j^2.$$

An analogous payoff expression applies also to bin k if the bid of the opponent exceeds the bid of the subject. Denote this payoff π_{kL}. The subjective probability of such outcome conditional on the bid of the opponent being in bin k is $1-q$. We then have the following result:

Proposition 2 *Suppose that $u(\cdot)$ is strictly concave and $u'(\cdot)$ is bounded away from 0 and ∞ on $[0, 20+s]$. Also suppose that the subjective belief is uniform. Then any optimal report $(r_1^*, .., r_5^*)$ satisfies:*

1. *if $k > 1$, then for any $i < k$ it holds that $r_i^* = r_L^*$; moreover, if $q < 1$, then $r_L^* < r_k^*$ and $r_L^* < 0.2$; if $q = 1$ and $k < 5$, then $r_L^* = r_k^* < 0.2$; if $q = 1$ and $k = 5$, then $r_L^* = r_k^* = 0.2$;*
2. *if $k < 5$, then for any $i > k$ it holds that $r_i^* = r_H$ with $r_H > r_k^*$ and $r_H > 0.2$.*

Proof of Proposition 2. The decision-maker solves

$$\begin{aligned} \max_{r_1,\dots,r_5\in[0,1]} \sum_{i<k} 0.2u\left(s + 10 + 20r_i - 10\sum_{j=1}^{5} r_j^2\right) + 0.2qu\left(s + 10 + 20r_k - 10\sum_{j=1}^{5} r_j^2\right) \\ + 0.2(1-q)u\left(10 + 20r_k - 10\sum_{j=1}^{5} r_j^2\right) + \sum_{i>k} 0.2u\left(10 + 20r_i - 10\sum_{j=1}^{5} r_j^2\right) \\ \text{s.t. } r_1 + \cdots + r_5 = 1. \end{aligned}$$

Following the same steps as in the proof of part 1 of Proposition 1, it follows that the probability bounds are never binding

in this problem and can therefore be ignored. Hence any solution $(r_1^*, \dots, r_5^*)$ to the unconstrained problem must satisfy the usual

first-order necessary conditions which, after canceling out a common multiplier 0.2×20, are given by

$$u'(\pi_i^*) - r_i^*\left[\sum_{j\neq k} u'(\pi_j^*) + qu'(\pi_{kH}^*) + (1-q)u'(\pi_{kL}^*)\right] = 0, \quad i \neq k,$$

$$\left[qu'(\pi_{kH}^*) + (1-q)u'(\pi_{kL}^*)\right] - r_k^*\left[\sum_{j\neq k} u'(\pi_j^*) + qu'(\pi_{kH}^*) + (1-q)u'(\pi_{kL}^*)\right] = 0,$$

where the asterisk denotes the resulting payoff if the corresponding state

is realized. Summing these 5 conditions gives

$$\left[\sum_{j\neq k} u'(\pi_j^*) + qu'(\pi_{kH}^*) + (1-q)u'(\pi_{kL}^*)\right] - \left(\sum_{j=1}^{5} r_j^*\right)\left[\sum_{j\neq k} u'(\pi_j^*) + qu'(\pi_{kH}^*) + (1-q)u'(\pi_{kL}^*)\right] = 0,$$

implying that $r_1^* + ... + r_5^* = 1$. As a result, any solution to the unconstrained problem also solves the constrained problem. The first-order

conditions also imply that

$$\frac{r_i^*}{u'(\pi_i^*)} = \frac{r_k^*}{qu'(\pi_{kH}^*) + (1-q)u'(\pi_{kL}^*)}, \quad i \neq k. \tag{4}$$

Since π_j is strictly increasing in r_j and $u(\cdot)$ is strictly

concave, it follows that: (1) if $k > 1$ and $q < 1$, then $r_i^* = r_L^*$ for any $i < k$ for some strictly positive (part 1 of Proposition 1) r_L^* and $r_L^* < r_k^*$; (2) if $k > 1$ and $q = 1$, then $r_i^* = r_L^*$ for

any $i \leq k$ for some strictly positive r_L^*; (3) if $k < 5$, then $r_i^* = r_H^*$ for any $i > k$ for some strictly positive r_H^* and $r_H^* > r_k^*$. The fact that $r_1^* + ... + r_5^* = 1$ then implies that: (1) if $k > 1$, then $r_L^* < 0.2$, unless $k = 5$ and $q = 1$, in which case $r_L^* = r_k^* = 0.2$; (2) if $k < 5$, then $r_H^* > 0.2$. ■

B Additional Analysis

In this appendix, we present a more in-depth analysis of responses of subjects who submitted at least one mistaken answer to the 1-through-10 control sub-question 3*c* (`wrong_simple_uniform = 1`) and the responses of those who did not submit a mistaken answer (`wrong_simple_uniform = 0`). For power reasons, we pool observations from all seven treatments.

Figure B1 presents the mean weight allocated to each of the five bins separately for the two groups. First. the figure readily shows that both groups show a considerable central bias on average. This is confirmed by a recomputation of Table 4 for each of the two groups. The results are qualitatively unchanged for either of the two groups, with the exception of the comparison of Bin 2 and Bin 3 for subjects with `wrong_simple_uniform= 1`. In this case, the Bonferroni-corrected p-values are 0.032 for the t-test and 0.004 for the signed rank test. Second, the figure also shows that the differences in the average weights between the two groups are small. This is confirmed by Table B1 that recomputes Table 1 separately for each of the two groups. Using the t-test, none of the differences in the mean weights are statistically significant at any conventional level even before the Bonferroni correction.

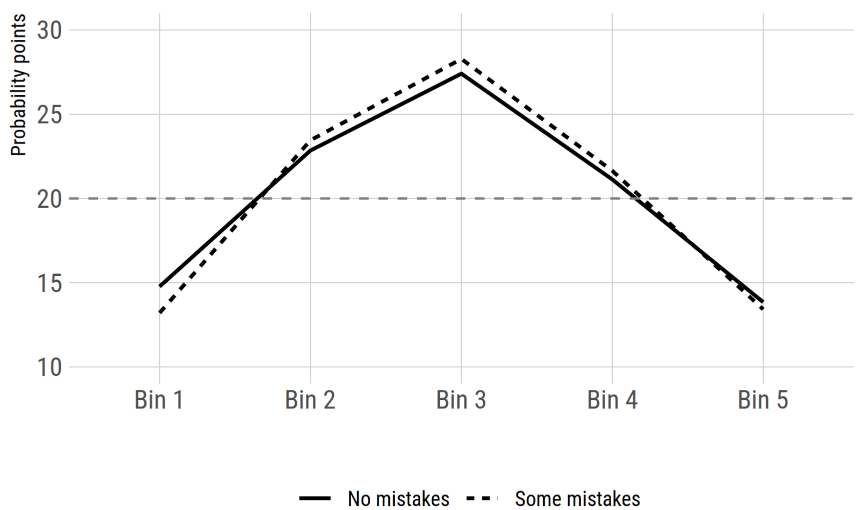

Figure B1: Mean weights within bins for subjects with no vs. some mistakes in answering the question 3c.

Table B2 presents a recomputation of Table 2 separately for each of the two groups. The "Non-Uniform" column emphasizes the incidence of uniform reporting between the two groups observed in the rightmost panel of Table 5. Using the t-test, the difference in the rate of non-uniform reporting is weakly statistically significant ($p = 0.089$). As for distances from the uniform report, the table shows that those with

Table B1: Mean weights within bins (additional analysis)

Subsample	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Subjects
`wrong_simple_uniform = 0`	14.77	22.85	27.41	21.13	13.85	272
`wrong_simple_uniform = 1`	13.21	23.45	28.28	21.64	13.42	107
t-statistic for difference	-1.27	0.43	0.58	0.43	-0.37	
Raw p-value	0.205	0.664	0.560	0.665	0.711	

Note: Means of the weights reported within each of the 5 bins are pooled across all treatments. The last two rows present heteroscedasticity-robust t-statistics and p-values for two-tailed tests of equality of the mean weights between the two groups.

Table B2: Non-uniform responses and distances from the uniform distribution (additional analysis)

Treatment	Non-Uniform	$d > 0.1$	$d > 0.2$	$d > 0.3$	Subjects
`wrong_simple_uniform = 0`	191 (70.2%)	178 (65.4%)	110 (40.4%)	68 (25.0%)	272
`wrong_simple_uniform = 1`	84 (78.5%)	78 (72.9%)	50 (46.7%)	16 (15.0%)	107
t-statistic for difference	1.71	1.44	1.11	-2.31	
Raw p-value	0.089	0.152	0.269	0.021	

Note: We list the number (and percentage) of subjects with non-uniform responses for the given subsample, pooled across all treatments. We also list the number (and percentage) of subjects whose reported beliefs deviate from the uniform distribution by at least a threshold sup-norm distance $d \in \{0.1, 0.2, 0.3\}$. The last two rows present heteroscedasticity-robust t-statistics and p-values for two-tailed tests of equality of percentages between the two groups.

`wrong_simple_uniform = 1`, although being somewhat more likely to deviate from the uniform report, deviate less from it in terms of distribution distance than do those with `wrong_simple_uniform = 0`. The differences in the rates of reporting with $d > 0.1$ and $d > 0.2$ are not statistically significant at any conventional level. The difference in the rates of reporting with $d > 0.3$ is statistically significant ($p = 0.021$), but in a direction suggesting that those with `wrong_simple_uniform = 1` are *less* likely to deviate far from the uniform report. However, after a Bonferroni correction none of the differences are statistically significant.

Overall, these results show that the CSP reporting is robust to the level of initial understanding of the uniform distribution as proxied by the response to the 1-through-10 sub-question 3c.

C Experimental instructions

C.1 General instructions

Initial screen:

You are about to participate in an experiment in which following the instructions carefully, making good decisions, and with a bit of luck, you can earn a considerable amount of money. Different participants may earn different amounts according to their choices. For your participation in the experiment you will earn an additional show-up fee of 2.5 Euro.

All the monetary values during the experiment are expressed in ECU (Experimental Currency Units). At the end of the experiment, the ECUs you earned will be converted into a cash payoff in Euro using the exchange rate 1 ECU = 20 euro cents and paid in cash privately.

New screen:

The experiment consists of 4 stages in the following order:

1. An Instruction Stage that we are currently going through. At the end of this stage you will be asked some control questions to verify your understanding of the task. After everybody answers correctly, we will proceed with the following stage.
2. A Decision Stage, in which you will make decisions and answer questions relevant towards your payoff.
3. A Demographic Questionnaire, in which you will be asked a few questions about your demographic and academic background.
4. A Feedback Stage, in which your earnings from the experiment will be determined and announced to you privately. You will not be given any feedback on the monetary outcome of your decisions before the Feedback Stage.

New screen:

You are about to participate in an auction against a hypothetical opponent. In this auction, there is a fictitious object for sale that you value at (60/80/100) ECUs. That is, if you win the auction, you obtain the object and receive (60/80/100) ECUs, but from this amount you have to subtract the price you will have to pay for the object.

Your task in this auction is to place a bid. This bid can be any integer number from 1 through (60/100) ECUs. A bid is a binding price offer to pay for the object in case you win the auction. Your opponent also places a bid that will be randomly drawn from the set of integers 1 through (60/100) ECUs. Each of these integers is equally likely to be drawn.

The winner of the auction is the bidder who places the highest bid. In case the two bids coincide, you are the winner. That is, you win if your bid is at least as high as your opponent's bid, otherwise you do not win. The winner pays the price for the object equal to his/her bid. That is, if you win, you receive (60/80/100) ECUs minus your bid. If you do not win, you do not receive anything and you do not pay anything either.

Control Questions: See below for details

Decision Stage: See below the treatment-specific instructions

Belief elicitation 1 stage:

New screen:

Before drawing your opponent's bid, we ask you to answer two short questions that allow you to earn some additional money. You will be paid for one of these two questions. At the end of the experiment, one of the participants will flip a coin to decide which question will be paid. You can earn up to 20 ECUs for the selected question.

Question 1. Please report your belief of your opponent's bid. We will provide five intervals. You are asked to report how likely you think your opponent's bid is to be in each of these intervals. The number in each input field you are asked to fill in is your percentage estimate of the likelihood of your opponent's bid being in that particular interval. The five percentages need to add up to 100. There will be an automatic checker to tell you what the current sum is as you enter the numbers.

You will be paid based on how closely your estimates match your opponent's bid. The exact formula (the so-called quadratic scoring rule) is complicated and the experimenters will be happy to explain it after the end of the experiment to those who are interested. However, in order to maximize your expected earnings from this procedure, you should report these likelihoods truthfully according to what you believe.

Here is some advice on how to fill in the five input fields:

- If you believe that your opponent's bid is more likely to be in a certain range, then assign higher percentages to intervals corresponding to that range and lower percentages to intervals corresponding to other ranges.

- If you believe that your opponent's bid is equally likely to be in several different intervals, then assign the same percentages to those intervals.
- Do not over-concentrate your assigned percentages in one or two intervals if you are not quite sure that your opponent's bid is in these intervals. Otherwise, if it turns out that your opponent's bid is in some other interval, you would earn little money from answering the question.
- On the other hand, do concentrate the entire 100 percent in one or two intervals if you feel confident that your opponent's bid is in this (these) interval(s). This will increase your earnings from answering this question.

New screen:

Here we repeat Question 1 for your convenience. Please provide your answer using the input fields below.

Question 1. Please report your belief of your opponent's bid. We will provide five intervals. You are asked to report how likely you think your opponent's bid is to be in each of these intervals. The number in each input field you are asked to fill in is your percentage estimate of the likelihood of your opponent's bid being in that particular interval. The five percentages need to add up to 100. There will be an automatic checker to tell you what the current sum is as you enter the numbers.

Belief elicitation 2 stage:

New screen:

Question 2. Please report your perceived probability of winning and of not winning the auction given the choice you made in the main task. The number in each box you are asked to fill in is your percentage estimate of the likelihood of that event. The two percentages need to add up to 100. There will be an automatic checker to tell you what the current sum is as you enter the numbers. You will be paid based on how closely your estimates match the outcome of the auction. The exact formula and the suggestions how to increase your payoffs are the same as in Question 1.

C.2 Control Questions

Please answer the following control questions. Answers to these questions are not relevant to your earnings. The computer will give you a feedback on whether your responses are correct or not. If you have any problems in answering, please raise your hand and an experimenter will come to assist you. After everyone answers correctly all the questions, we will proceed with the decision stage.

1. Suppose that your opponent bids 25 ECUs.

    a. If you bid 21 ECUs, how much will you earn in ECU?

    b. If you bid 38 ECUs, how much will you earn in ECU?

    c. If you bid 62 ECUs, how much will you earn in ECU?

    d. If you bid 79 ECUs, how much will you earn in ECU?

2. Now suppose that your opponent bids 75 ECUs.

    a. If you bid 21 ECUs, how much will you earn in ECU?

    b. If you bid 38 ECUs, how much will you earn in ECU?

    c. If you bid 62 ECUs, how much will you earn in ECU?

    d. If you bid 79 ECUs, how much will you earn in ECU?

3. What is the probability of your opponent's bid being in the range:

    a. 39 through 72

    b. 22 though 47

    c. 1 through 10

    d. 16 through 56

    e. 62 through 100

4. Your opponent's bid depends on your bid. YES/NO

5. Suppose the bid of your opponent is identical to your bid. Will you earn a positive amount? YES/NO

Please note that the numbers used in these questions are for illustrative purposes only. They are not meant to be a guidance for your choice.

C.3 Treatment-specific instructions

C.3.1 Auction 100/100 with visualization treatment

To help you visualize your decision, on your display you will see a square composed of 100 boxes numbered 1 through 100. One of these boxes corresponds to your opponent's bid. You do not know which one it is, however. You only know that it can be any of the 100 boxes with equal probability.

You initiate the bidding process by first entering your intended decision into an input field and clicking on "Evaluate." At this point, the originally grey boxes change color. The boxes turning blue are those whose number is less than or equal to your bid. The boxes turning yellow are those whose number exceeds your bid. If the opponent's bid corresponds to one of the blue boxes, you will earn the difference between 100 ECUs and your bid. If the opponent's bid corresponds to one of the yellow boxes, your will earn zero. You are free to evaluate different bids in this way.

When you are confident about your choice, submit it by clicking on the "Submit" button and then reconfirm it by clicking on "Confirm".

At the end of the experiment, after answering some questions and filling out a short questionnaire, we will randomly determine the number of the box corresponding to your opponent's bid. This will be done by one of the participants randomly drawing a token from a bag containing 100 tokens numbered 1 through 100.

New screen:

Please choose your bid. Using the input field and the "Evaluate" button, you are free to evaluate as many different bids as you wish. The boxes turning blue are those whose number is less than or equal to your bid. The boxes turning yellow are those whose number exceeds your bid. If the opponent's bid corresponds to one of the blue boxes, you will earn the difference between 100 ECUs and your bid. If the opponent's bid corresponds to one of the yellow boxes, your will earn zero. When you are ready to submit your final decision, click on the "Submit" button and then confirm your choice by clicking on "Confirm".

C.3.2 Auction 100/100 without visualization treatment

Identical to Auction 100/100 with visualization treatment but without visualization or mentions of visualization.

C.3.3 Bomb Risk Elicitation Task treatment

Your task is to decide on the number of boxes to collect out of 100 such boxes numbered 1 through 100. You collect the boxes starting from box number 1, continuing until the box whose number is equal to the number of boxes you decide to collect. Exactly one of these 100 boxes contains a bomb. You do not know the bomb's location. You only know that it is equally likely to be in any of the 100 boxes.

If the number of the box in which the bomb is located is higher than the number of boxes you collected, you do not collect the bomb and you earn 1 ECU for each collected box. If the number of the box in which the bomb is located is lower than or equal to the number of boxes you collected, you do collect the bomb and you earn zero.

New screen:

To help you visualize your decision, on your display you will see a square composed of 100 numbered boxes. There is a bomb in one of these boxes. You do not know which one it is, however. You only know that it can be in any of the 100 boxes with equal probability.

You initiate the decision process by first entering your intended decision into an input field and clicking on "Evaluate." At this point, the originally grey boxes change color. The boxes turning yellow are those that you are deciding to collect. The boxes turning blue are those you are deciding to not collect. If the bomb is in one of the yellow boxes, you will earn zero. If the bomb is in one of the blue boxes, you will earn 1 ECU for every (yellow) box collected. You are free to evaluate different numbers of collected boxes in this way.

At the end of the experiment, after answering some questions and filling out a short questionnaire, we will randomly determine the number of the box containing the bomb. This will be done by one of the participants randomly drawing a token from a bag containing 100 tokens numbered 1 through 100.

Instructions from Questions 1 and 2:

Identical to the Auction 100/100 with visualization treatment but "your opponent's bid" was replaced by "the position of the bomb" and "winning and of not winning the auction" was replaced by "collecting and not collecting the bomb."

C.3.4 Auction 80/100 treatment

Identical to Auction 100/100 with visualization treatment but the value to the subject was 80.

Added to first screen in Auction 100/100 with visualization treatment:

Note that if you bid more than 80 you make losses in case you win.

Added to second screen in Auction 100/100 with visualization treatment:

In case you bid more than 80 the boxes in excess turn red signaling that you would make losses if you win.

C.3.5 Auction 60/100 treatment

Identical to Auction 80/100 with visualization treatment but the value to the subject was 60 and the warnings were for bids greater than 60.

C.3.6 Auction 60/60 treatment

In this treatment both the value of the subject and of the opponent are set to 60.

Subjects bid choosing a number from 0 to 60, and report beliefs of the opponent's bid which is uniformly distributed in the interval 0-60.

New screen:

You are about to participate in an auction against a hypothetical opponent. In this auction, there is a fictitious object for sale that you value at 60 ECUs. That is, if you win the auction, you obtain the object and receive 60 ECUs, but from this amount you have to subtract the price you will have to pay for the object.

Your task in this auction is to place a bid. This bid can be any integer number from 1 through 60 ECUs. A bid is a binding price offer to pay for the object in case you win the auction. Your opponent also places a bid that will be randomly drawn from the set of integers 1 through 60 ECUs. Each of these integers is equally likely to be drawn.

The winner of the auction is the bidder who places the highest bid. In case the two bids coincide, you are the winner. That is, you win if your bid is at least as high as your opponent's bid, otherwise you do not win. The winner pays the price for the object equal to his/her bid. That is, if you win, you receive 60 ECUs minus your bid. If you do not win, you do not receive anything and you do not pay anything either.

New screen:

To help you visualize your decision, on your display you will see a rectangle composed of 60 boxes numbered 1 through 60. One of these boxes corresponds to your opponent's bid. You do not know which one it is, however. You only know that it can be any of the 60 boxes with equal probability.

You initiate the bidding process by first entering your intended decision into an input field and clicking on "Evaluate." At this point, the originally grey boxes change color. The boxes turning blue are those whose number is less than or equal to your bid. The boxes turning yellow are those whose number exceeds your bid. If the opponent's bid corresponds to one of the blue boxes, you will earn the difference between 60 ECUs and your bid. If the opponent's bid corresponds to one of the yellow boxes, you will earn zero. You are free to evaluate different bids in this way.

At the end of the experiment, after answering some questions and filling out a short questionnaire, we will randomly determine the number of the box corresponding to your opponent's bid. This will be done by one of the participants randomly drawing a token from a bag containing 60 tokens numbered 1 through 60.

C.3.7 Auction 60/60 expand treatment

New screen:

You are about to participate in an auction against a hypothetical opponent. In this auction, there is a fictitious object for sale that you value at 60 ECUs. That is, if you win the auction, you obtain the object and receive 60 ECUs, but from this amount you have to subtract the price you will have to pay for the object.

Your task in this auction is to place a bid. This bid can be any integer number from 1 through 60 ECUs. A bid is a binding price offer to pay for the object in case you win the auction. Your opponent also places a bid that will be randomly drawn from the set of integers 1 through 60 ECUs. Each of these integers is equally likely to be drawn.

The winner of the auction is the bidder who places the highest bid. In case the two bids coincide, you are the winner. That is, you win if your bid is at least as high as your opponent's bid, otherwise you do not win. The winner pays the price for the object equal to his/her bid. That is, if you win, you receive 60 ECUs minus your bid. Note that if you bid more than 60 you make losses in case you win. If you do not win, you do not receive anything and you do not pay anything either.

New screen:

To help you visualize your decision, on your display you will see a square composed of 100 boxes numbered 1 through 100. One of these boxes corresponds to your opponent's bid. You do not know which one it is, however. You only know that it can be any of the 60 boxes with equal probability.

You initiate the bidding process by first entering your intended decision into an input field and clicking on ”Evaluate.” At this point, the originally grey boxes change color. The boxes turning blue are those whose number is less than or equal to your bid. In case you bid more than 60 the boxes in excess turn red signaling that you would make losses if you win. The boxes turning yellow are those whose number exceeds your bid. If the opponent’s bid corresponds to one of the blue or red boxes, you will earn the difference between 60 ECUs and your bid. If the opponent’s bid corresponds to one of the yellow boxes, your will earn zero. You are free to evaluate different bids in this way.

At the end of the experiment, after answering some questions and filling out a short questionnaire, we will randomly determine the number of the box corresponding to your opponent’s bid. This will be done by one of the participants randomly drawing a token from a bag containing 60 tokens numbered 1 through 60.

D Screenshots

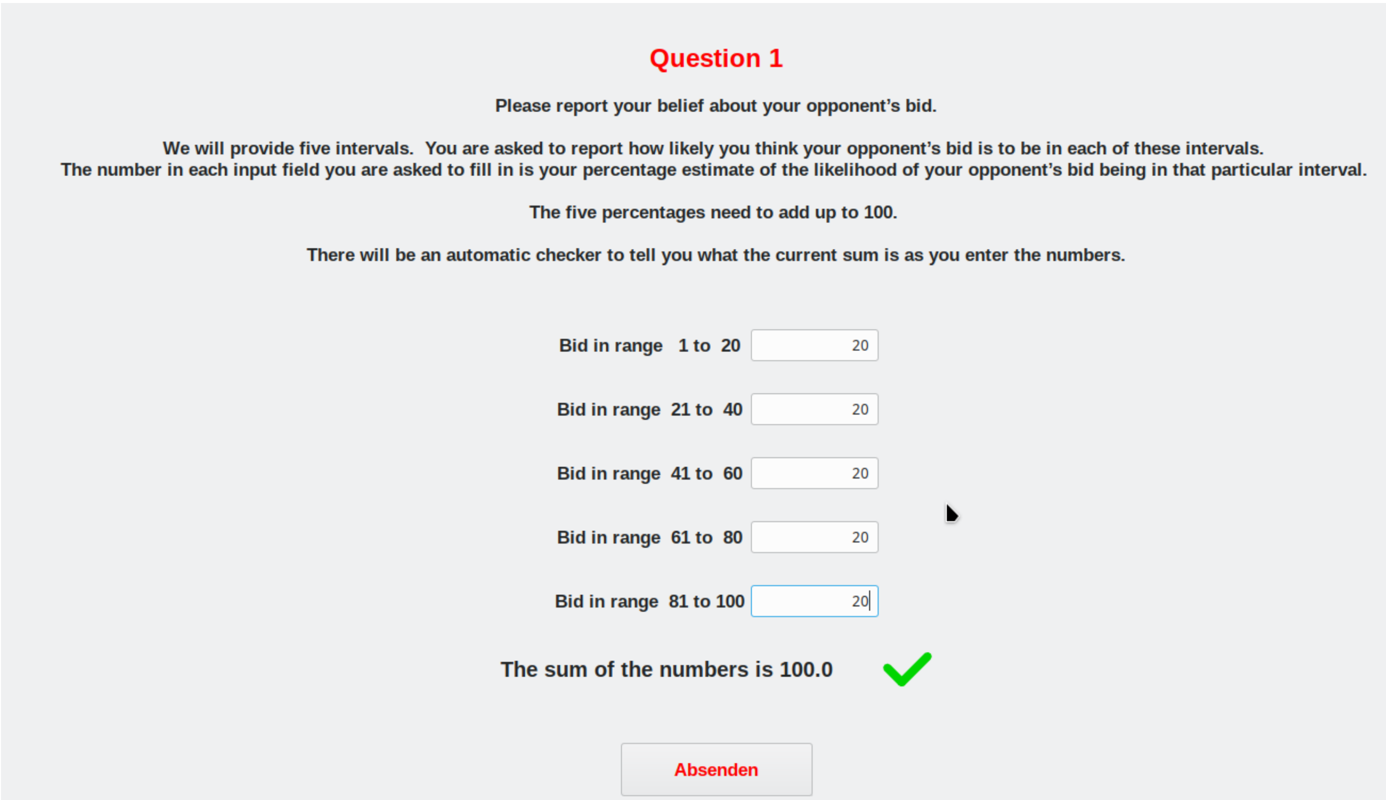

Figure D2: Screenshot of the belief elicitation of the random-draw opponent's strategy after a subject attempts to report weights that correctly sum to 100.

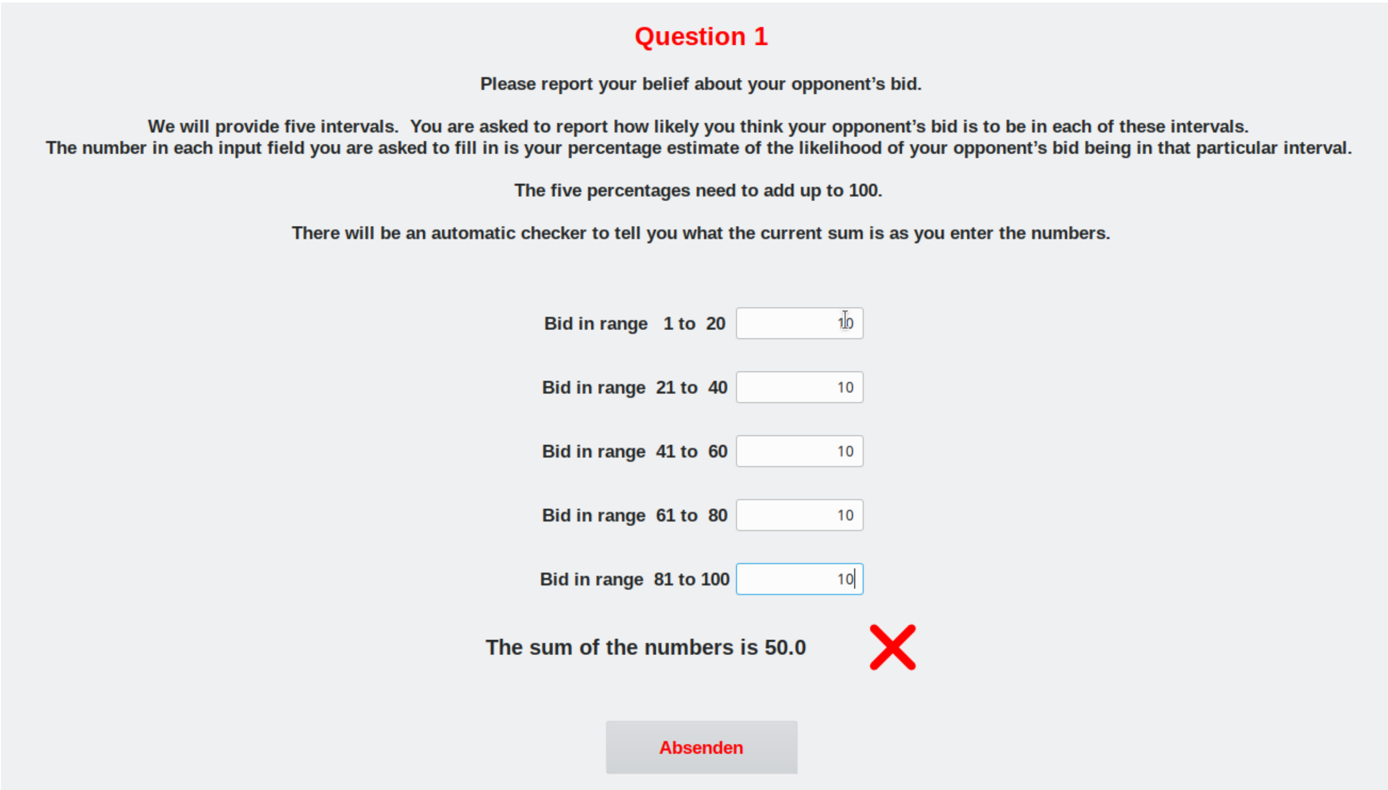

Figure D3: Screenshot of the belief elicitation of the random-draw opponent's strategy after a subject attempts to report weights that do not sum to 100.

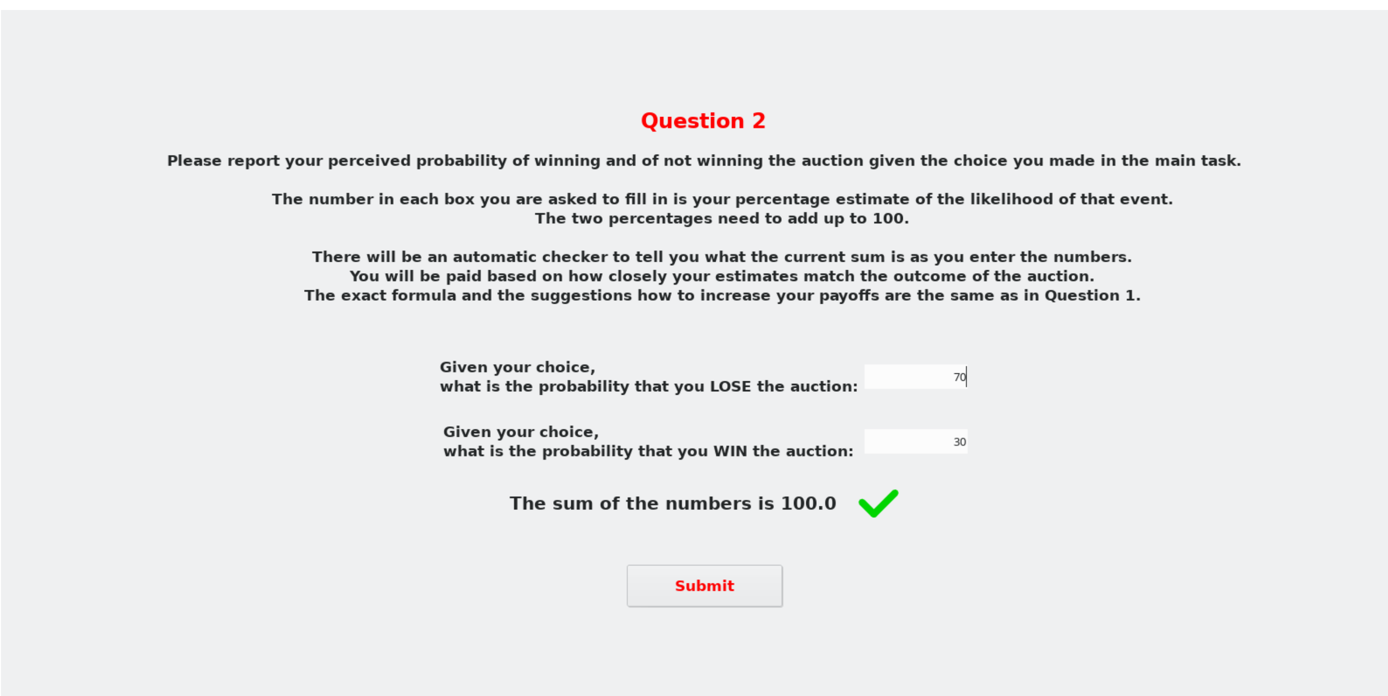

Figure D4: Screenshot of the belief elicitation of the probability of winning the auction after a subject attempts to report weights that correctly sum to 100.

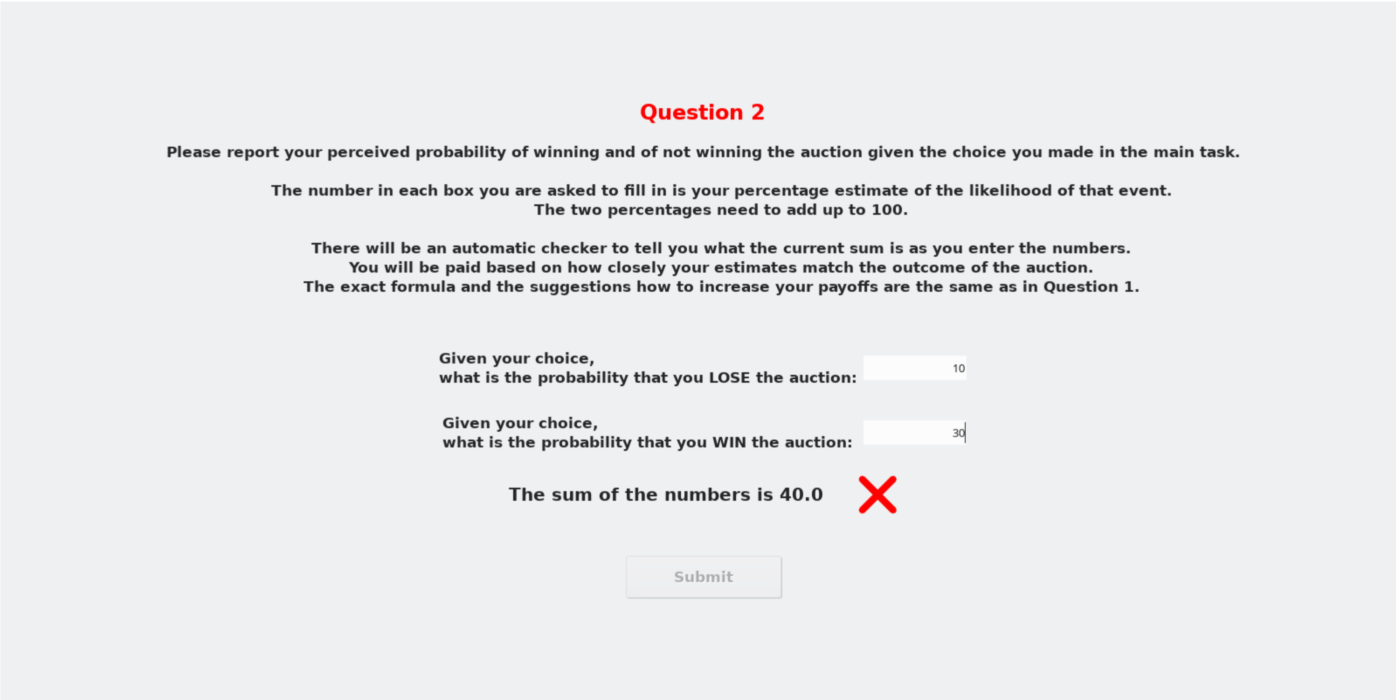

Figure D5: Screenshot of the belief elicitation of the probability of winning the auction after a subject attempts to report weights that do not sum to 100.

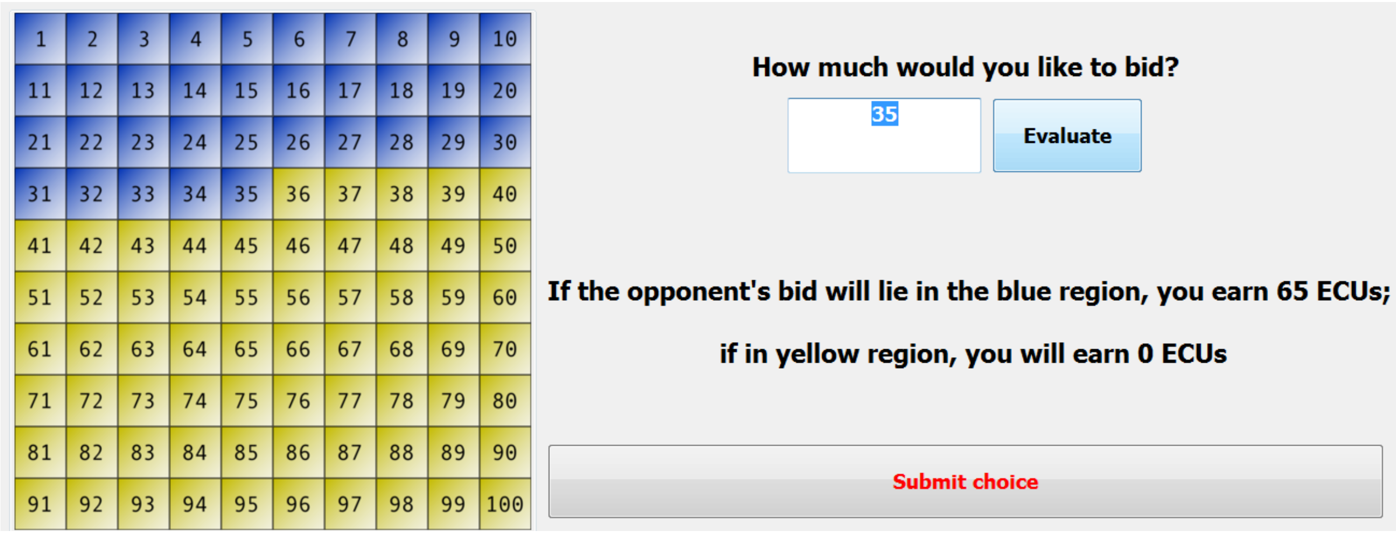

Figure D6: Screenshot of the bid submission screen in auction 100/100.

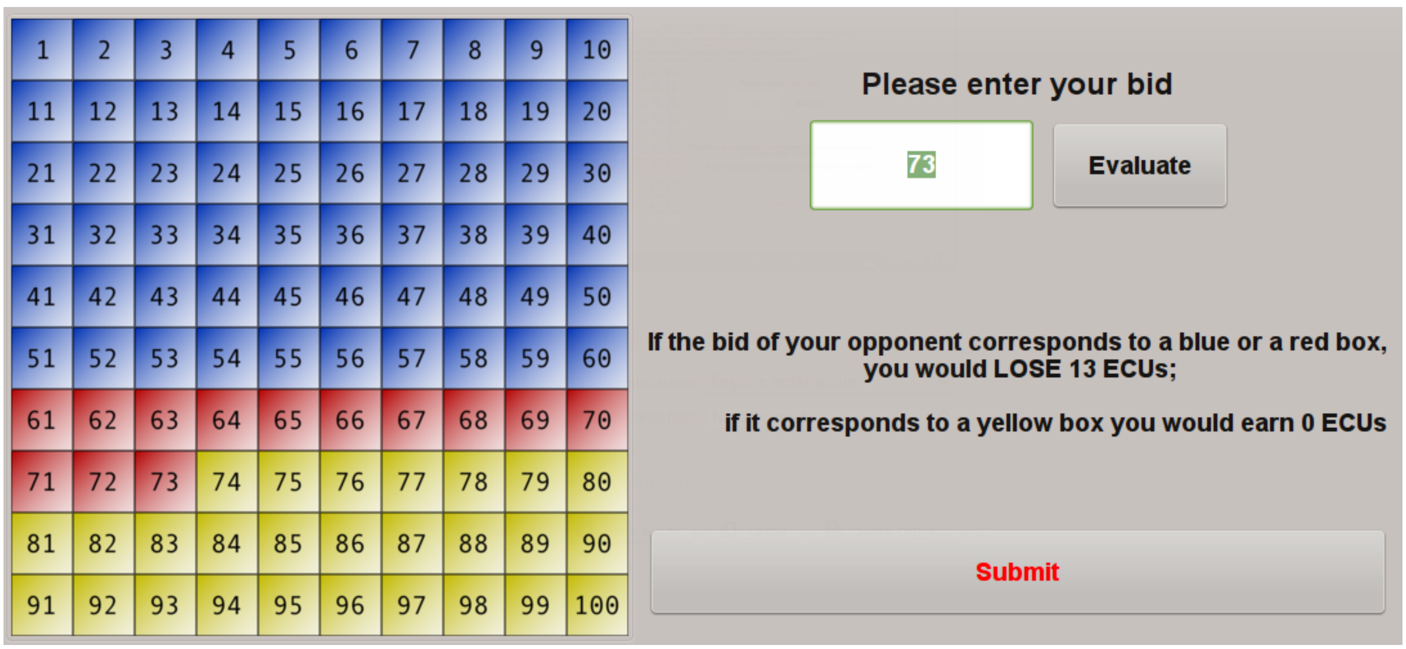

Figure D7: Screenshot of the bid submission screen in auction 60/100.