



## Research paper

What drives demand for loot boxes? An experimental study<sup>☆</sup>Simon Cordes<sup>a</sup>, Markus Dertwinkel-Kalt<sup>b,c,\*</sup>, Tobias Werner<sup>d,e</sup><sup>a</sup> University of Bonn, Germany<sup>b</sup> University of Münster, Germany<sup>c</sup> Max Planck Institute for Research on Collective Goods, Germany<sup>d</sup> Center for Humans and Machines at the Max Planck Institute for Human Development, Germany<sup>e</sup> Düsseldorf Institute for Competition Economics, Germany

## ARTICLE INFO

## Keywords:

Gaming

Video games

Gambling

Loot boxes

Microtransactions

## ABSTRACT

The market for video games is booming, with in-game purchases accounting for a substantial share of developers' revenues. Policymakers and the general public alike are concerned that so-called "loot boxes" – lotteries that offer random rewards to be used in-game – induce consumers to overspend on video games. We provide experimental evidence suggesting that common design features of loot boxes (such as opaque odds and positively selected feedback) indeed induce overspending by inflating the belief of winning a prize. In combination, these features *double* the average willingness-to-pay for lotteries. Based on our findings, we argue for the need to regulate the design of loot boxes to protect consumers from overspending.

## 1. Introduction

The market for (online) video games has been booming in recent years, with in-game purchases accounting for a substantial share of developers' revenues. In 2020 alone, so-called "loot boxes" generated \$15 billion of worldwide revenue, and projections suggest that 230 million people will spend money on loot boxes by 2025.<sup>1</sup> Loot boxes are digital lotteries in video games that – similar to gambling – offer *random* rewards to be used in-game. While loot boxes share many similarities with gambling (Drummond and Sauer, 2018), surprisingly little regulation is in place that would restrict their design and the way they are priced.<sup>2</sup> At the same time, policymakers and the general public alike are concerned that loot boxes induce consumers – in particular, those susceptible to gambling – to overspend on video games.<sup>3</sup>

This concern is amplified by the fact that loot boxes are designed and marketed in ways that obfuscate the chances of winning different rewards. First, the odds are often censored. As a specific example, consider the football simulation *FIFA Ultimate Team*, where gamers build a team of players that vary in strength. Gamers can buy packs that offer lotteries over players. The odds,

<sup>☆</sup> We thank Mats Köster for his sustained support during the project. Furthermore, we thank Daniel Evans, Jonas Frey, Valentin Kecht, Hans-Theo Normann, Hannah Schildberg-Hörsch, Georg Schneider, Vasilisa Werner, and Florian Zimmermann for helpful comments. We gratefully acknowledge financial support by the German Research Foundation (DFG project No. 462020252).

\* Corresponding author at: University of Münster, Germany.

E-mail addresses: [simon.cordes@uni-bonn.de](mailto:simon.cordes@uni-bonn.de) (S. Cordes), [markus.dertwinkel-kalt@wiwi.uni-muenster.de](mailto:markus.dertwinkel-kalt@wiwi.uni-muenster.de) (M. Dertwinkel-Kalt), [werner@mpib-berlin.mpg.de](mailto:werner@mpib-berlin.mpg.de) (T. Werner).

<sup>1</sup> See, for instance, [https://www.juniperresearch.com/press/video-game-loot-boxes-to-generate-over-\\$20-billion](https://www.juniperresearch.com/press/video-game-loot-boxes-to-generate-over-$20-billion) (accessed on September 16th, 2022).

<sup>2</sup> Recently, 20 consumer organizations from 18 European countries have suggested that loot boxes should be classified as gambling and therefore regulated (The Norwegian Consumer Council, 2022). Additionally, the Federal Trade Commission (FTC) is investigating loot boxes following concerns from U.S. legislators that they may be similar to gambling (Federal Trade Commission, 2020).

<sup>3</sup> See, for instance, <https://www.theguardian.com/society/2021/apr/02/video-game-loot-boxes-problem-gambling-betting-children> (accessed on September 16th, 2022).

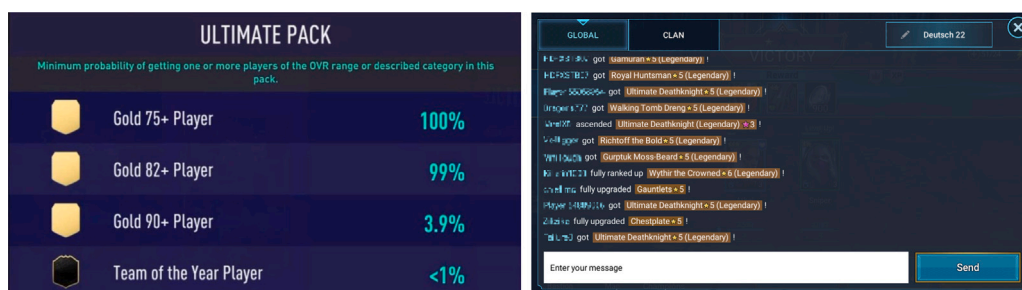


Fig. 1. Typical design features, censored odds (left) and selected feedback (right), of loot boxes.

however, are provided, if at all, only for a coarse set of intervals, bunching together players of very different strengths (see the left panel of Fig. 1). At the extreme, the worst player in an interval is around 1000 times less valuable than the best player. The Norwegian Consumer Council (2022) argues that gamers, therefore, overestimate the value of these lotteries. Moreover, the FTC asserts that false or inaccurate odds may violate the Federal Trade Commission Act, specifically Section 5, which prohibits firms' unfair or deceptive practices (Federal Trade Commission, 2020).

Second, gamers often receive highly selective feedback on the rewards other gamers have obtained. In the mobile game *Raid: Shadow Legends*,<sup>4</sup> for example, gamers receive a notification whenever another player wins a rare reward (see the right panel of Fig. 1). This feature leads to a constant but selected stream of signals about rewards from loot boxes. As only rare rewards are reported, this provides them with a biased sample of the reward distribution.

Going further, game developers not only pay content providers (on *Youtube* or *Twitch*) to open loot boxes on their shows, but they allegedly also offer them better odds.<sup>5</sup> According to The Norwegian Consumer Council (2022, p.44), observing such a biased sample of the reward distribution “reinforces the players’ belief that they might be similarly lucky”. While one could easily imagine that both features contribute to overspending on loot boxes, there is a lack of systematic evidence supporting this claim.

In order to fill this gap, we experimentally investigate what drives the willingness-to-pay for loot boxes. In a between-subject design, we focus on the effects of censoring the odds and providing gamers with a selected sample of the reward distribution. We do so because these design features arguably do not provide any utility to gamers, while their sole purpose seems to be making consumers overspend on loot boxes. Indeed, we find evidence that both features substantially increase the willingness-to-pay for lotteries. Censoring the odds of a lottery increases a subject’s willingness-to-pay compared to a baseline treatment. Also, simply providing subjects with a selective sample of the reward distribution increases their willingness-to-pay. Combining censored odds with a selected sample increases the willingness-to-pay by 100%. In a between-subjects design, we demonstrate that both features – censoring and sampling – increase the subject’s willingness-to-pay by inflating her belief of winning a high reward. It is consistent with existing work that, for instance, shows that selected feedback affects economic behavior via a belief channel (e.g., Barron et al., 2024). Overall, our results suggest that the design of loot boxes – combining censored odds with selected feedback – contributes to overspending in the video game market and thus supports a case for regulating loot boxes.

We introduce our experimental design in Section 2. Subjects repeatedly state their willingness-to-pay (WTP) for different monetary lotteries with three potential prizes, one of which is zero. In a *Control* condition, we transparently describe the odds of the lotteries and do not provide additional information to the subjects. We assume that this *Control* condition identifies a subject’s true WTP, and define overspending relative to this benchmark. We implement three treatments that capture the features of loot boxes discussed above. In *Censored*, subjects only learn the total probability of winning a non-zero prize, but not the exact probability of winning the highest prize. In *Sample*, we provide subjects with the full prize distribution and a selected sample thereof; that is, they observe the five highest outcomes in a sample of 400 draws. Finally, *Joint* combines both: subjects observe the censored prize distribution and a selected sample thereof. This last treatment resembles the current design of loot boxes most closely. Notably, our experimental design eliminates all features of loot boxes that may provide utility beyond winning a reward, such as a nice design or visual effects. Instead, we isolate the features of loot boxes that almost certainly do not affect a gamer’s material utility and can thus be interpreted as inducing mistakes.

Section 3 presents our main results. Compared to the *Control* condition, the average WTP increases by roughly 45% in *Censored* and *Sample*, respectively. The average WTP doubles in *Joint* compared to the control condition. The subjects’ beliefs show a similar pattern as the WTP across the different treatment conditions. Moreover, we identify the beliefs as the main channel of our treatment effects. Once we control for stated beliefs (i.e., the mediator), the average WTPs in *Censored*, *Sample*, and *Joint* do not differ significantly from that in *Control* anymore. It demonstrates that censored odds and selected feedback increase the subjects’ WTPs by inflating their beliefs of winning a high prize.

<sup>4</sup> In 2022, three years after its release, the game surpassed \$1bn in lifetime revenue. For details, see <https://gamingonphone.com/news/raid-shadow-legends-surpasses-1-billion-in-lifetime-revenue/> (accessed on January 2nd, 2023).

<sup>5</sup> See, for instance, <https://gamerant.com/ftc-loot-boxes-better-odds-sponsored-streamers/> (accessed on December 22nd, 2022).

In Section 4, we provide additional evidence on the underlying mechanism as well as for the relevance of our results. First, we study correlates of a survey measure on loot-box overspending. Consistent with the prior literature (e.g., Zendle and Cairns, 2018), survey measures of gambling behavior correlate with overspending on loot boxes. Controlling for these measures of gambling behavior, we still find a positive association between the average WTP for the lotteries in our experiment and survey measures of overspending on loot boxes. This speaks to the external validity of our findings and supports our view that our experimental measure of gambling behavior is informative for real-world overspending on loot boxes.

Second, we restrict the sample to decisions for which stated beliefs are “realistic” in that they are consistent with the provided information. Here, we find a precisely estimated zero difference in average beliefs between *Censored* and *Joint*. In either case, subjects tend to assign equal probabilities to the non-zero prizes. This is consistent with evidence on people naively applying a “50-50 heuristic” when being uncertain about the problem they face (e.g., Fischhoff and Bruine De Bruin, 1999).

Third, we ran a robustness experiment to address the concern that the lotteries are offered by the experimenter, not a firm, trying to maximize profits. The treatment *Info* replicates *Joint* but adds unbiased information on the reward distribution and explicitly tells the subject that the odds are not 50-50. While the additional information makes subjects less optimistic about winning the highest prize, it does not affect the average WTP. Our findings highlight the robustness of our main treatments and show that even in the presence of further unbiased information, the design features of loot boxes promote overspending.

We conclude in Section 5 by discussing tentative policy implications and challenges in their implementation. Our results highlight the complexities of regulating loot boxes. Current regulatory efforts, like those in Germany, focus on labeling video games that contain loot boxes. This may fall short as such a regulation does not address the core issue of loot box design, which we find to cause overspending. Additionally, our findings extend beyond digital loot boxes to other markets, such as trading cards and online gambling, where similar regulatory measures could help to reduce overspending.

**Related literature.** First, we contribute to the literature on gaming, specifically loot boxes. A series of papers have established a positive correlation between survey measures of gambling and overspending on loot boxes (Drummond and Sauer, 2018; Zendle and Cairns, 2018; Drummond et al., 2020). We strengthen this link by providing causal evidence on how key design features of loot boxes affect the WTP for lotteries. Chen et al. (2021) develop a model of optimal loot box pricing, assuming gamers maximize expected utility (EU). Gan (2022) shows that for buyers that have prospect theory preferences and are naive about them, selling a product via a loot box (that delivers the good with some constant probability in each period) represents the seller’s uniquely optimal selling mechanism. Complementary to this work, we single out those features of loot boxes that are irrelevant for EU-maximizers but inflate demand by behavioral consumers—censoring and sampling.

The fact that game developers introduce “ambiguity” by censoring the odds of loot boxes connects our paper to the behavioral literature on choice under risk and ambiguity. Actually, firms should shy away from introducing payoff ambiguity as people generally have a strong distaste for ambiguity (see the large literature on ambiguity aversion building on Gilboa and Schmeidler, 1989; Schmeidler, 1989). If people are averse to ambiguity, ambiguity in payoffs should lower their willingness-to-pay. In our context, however, consumers appear to be ambiguity seeking as ambiguity increases their willingness-to-pay.

This observation aligns with recent literature that has challenged the universality of ambiguity aversion (e.g., Trautmann and Van De Kuilen, 2015; Kocher et al., 2018), finding that people tend to seek ambiguity when winning probabilities are small (see, e.g., Dimmock et al., 2016; Chandrasekher et al., 2022, and references therein). We enrich the existing literature by examining (i) a specific form of ambiguity that emerges from censoring the upper tail of a probability distribution, which firms can utilize to increase profits, and (ii) its application in the important domain of gaming, where firms can exploit it particularly effectively.

Notably, not only ambiguous or “vague” probabilities, but also vague outcomes have been investigated in the literature: subjects are vagueness-seeking if vagueness regards outcomes (Du and Budesu, 2005). In our loot box context, however, only vagueness in probabilities plays a role.

Further, our work relates to the literature studying real-world state lottery choice (e.g., Clotfelter and Cook, 1990; Rogers and Webley, 2001; Lockwood et al., 2024; Kachurka et al., 2021). In contrast to this strand of literature, we focus on an application to loot box design – a lottery design that uses censoring and selected sampling – methods typically not observed in real-world state lotteries. Further, the odds of winning in real-world lotteries are typically available—unlike in our setting.

Finally, we add to the literature on biased inferences from (non-)disclosed data. Empirical evidence from the lab and field suggests that individuals often draw wrong inferences from selectively disclosed data in strategic settings (Bolton et al., 2007; Koehler and Mercer, 2009; Brown et al., 2012; Benndorf et al., 2015; Deversi et al., 2021; Jin et al., 2021, 2022) and non-strategic ones (Esponda and Vespa, 2018; Barron et al., 2024; Enke, 2020; López-Pérez et al., 2022). While the former work relates to the Censored condition, the latter work resembles the Sample condition. We find that subjects naively bias censored probabilities towards a uniform distribution. By closely resembling common features of loot boxes, our design allows us to provide more nuanced insights into how to regulate their design.

## 2. Experimental design

### 2.1. Experimental setup

We develop an experimental design that allows us to test for the effects of two key features of loot boxes on the WTP for lotteries. First, the displayed odds of loot boxes are often censored. Second, gamers typically receive (positively) selected feedback on the reward distribution. We implement three treatments and one control condition to identify the effect of each feature in isolation – i.e., the effect of censored odds and selected feedback – as well as how both features interact with each other.

All subjects sequentially state their WTP for five lotteries.<sup>6</sup> Each lottery pays a non-zero prize with probability  $q\%$  and zero otherwise. The non-zero prize is either 10 Coins or  $x$  Coins. Both probabilities and prizes vary across decisions: in each decision, we independently draw (without replacement) a probability  $q \in \{10, 20, 30, 40, 50\}$  and a prize  $x \in \{100, 120, 140, 160, 180\}$ . Probability and prize pairs are drawn at the subject level, so that different subjects may observe different lotteries in a different order. The high prize of  $x$  Coins is always realized with probability 1%. For example, if  $q = 10\%$  and  $x = 100$ , the lottery pays 0 Coins with probability 90%, 10 Coins with probability 9%, or 100 Coins with probability 1%. Before stating their WTP for a lottery, we ask subjects to state their belief on how often they would win this high prize in 100 draws (see Fig. 2(a) for a screenshot and Section 2.2 for an interpretation).

In a between-subject design, we vary the amount of information that subjects receive on the lotteries across four conditions. We next describe these four conditions in more detail. Screenshots of the instructions and decision screens can be found in Online Appendix C.

**Control.** In the control condition, subjects learn the full probability distribution of the different lotteries. More specifically, as illustrated in Fig. 2(a), they learn the exact probability of winning 10 Coins and  $x$  Coins, respectively. Subjects do not get any additional information on the lotteries.

**Censored.** In the treatment *Censored*, subjects observe censored versions of the lotteries. For example in Fig. 2(b), they only learn the probability of receiving a non-zero prize, but not the exact probabilities of receiving 10 Coins or  $x = 100$  Coins, respectively. This mimics the censoring strategies of video game designers such as *EA Sports* who do not provide gamers with the full probability distribution (see the left panel of Fig. 1). Other than in *Control*, to assess the value of a lottery, subjects have to form a belief about the probability of receiving  $x = 100$  Coins. Based on existing research (e.g., Fischhoff and Bruine De Bruin, 1999), we expect subjects to overestimate this probability, likely biasing it towards 50%.

**Sample.** In the treatment *Sample*, subjects again learn the full probability distribution of each lottery, but on top, observe a sample from this distribution. As illustrated in Fig. 2(c), we present subjects the 5 highest outcomes in a sample of 400 actual draws from the lottery drawn at the subject level.<sup>7</sup> Notably, subjects receive transparent information on how the outcomes are chosen. This treatment is motivated by the common practice in video games of announcing rare prizes other players have obtained (see the right panel of Fig. 1). Importantly, because subjects observe the full probability distribution, the sample does not contain any new information regarding the value of the lottery. Still, existing research (e.g., Barron et al., 2024) suggests that observing a series of high draws from a distribution may increase a subject's WTP.

**Joint.** The treatment *Joint* combines both of the above: subjects observe a censored version of the lotteries together with the 5 highest outcomes in a sample of 400 draws. Unlike in *Sample*, the sample does contain information about the underlying probability distribution in this case. With censored odds, subjects arguably overestimate the low probability of winning  $x$  Coins initially. Then, if all subjects were Bayesian, the average belief upon observing the sample should decrease, moving closer to the truth. If, in contrast, subjects naively infer from a series of good draws that the lottery has to be even better than they initially thought, the average belief upon observing the sample should go up. Hence, compared to *Censored*, also the average WTP should increase.

**Discussion of the design.** Our design aims to contain the most essential and generalizable aspects of loot boxes while keeping the design sufficiently simple for subjects. We thus omit some aspects that are worth discussing. Firstly, firms that offer loot boxes are incentivized to choose the odds to maximize their profit. We deliberately omitted this feature in our setup. Recent experimental evidence by Deversi et al. (2021) and Jin et al. (2021) demonstrates that subjects often do not account for the strategic disclosure motives of other parties in sender–receiver games. Subjects lack skepticism regarding undisclosed information, even if these motives are transparently communicated and repeated interaction occurs. Therefore, it is plausible that video gamers do not consider the profit motive when evaluating the associated probabilities and purchasing loot boxes. Additionally, video gaming is a leisure activity for most people, and the firm's incentives are likely not at the top of their minds. Secondly, unlike most video games, we chose to be explicit about the selection process in the *Sample* condition. It allows us to keep beliefs about the selection process constant. If anything, this should attenuate treatment effects, as participants should be less optimistic in the *Sample* condition compared to the real world. Consequently, we identify a lower bound treatment effect in the *Sample* condition. To further scrutinize the argument that those differences to real-world loot boxes do not invalidate our experimental design, we conduct an additional treatment discussed in Section 4.3 that provides further information to the subjects. It highlights the generalizability and robustness of our experimental design.

## 2.2. Conceptual framework

We sketch a simple model that motivates our experimental design as well as our analysis of the experimental data. Consider a lottery  $Z$  with a distribution  $G^*$  over prizes (such as player cards in *FIFA Ultimate Team*). The lottery is presented in a “format”  $f$  (such as our different treatments), which captures the description of the odds or the feedback provided to gamers. We assume that gamers form a subjective belief  $G_f$  over the prize distribution that depends on the format.

<sup>6</sup> Collecting several decisions within-subject has many advantages. First, it allows us to reduce individual-level decision noise that might be caused by factors orthogonal to the lottery design. Second, it increases statistical power, enabling us to measure treatment effects more precisely. Third, specific characteristics of one lottery might systematically distort participants' behavior. Using a large set of lotteries allows us to rule out lottery-level effects. Reassuringly, we find no order effects in stated WTP as shown in Figure A.1.

<sup>7</sup> For each subject, samples were drawn independently.

| Lottery     |         |          |           |
|-------------|---------|----------|-----------|
| Probability | 90 %    | 9 %      | 1 %       |
| Payoff      | 0 coins | 10 coins | 100 coins |

Imagine you would play the lottery **100 times**. How often do you think would you win **100 coins**?

0 times  100 times

I believe that I would win 100 coins exactly **1 time** if I would play the lottery 100 times.

How much are you willing to pay for this lottery?

Coins

a) Control

| Lottery     |         |                              |
|-------------|---------|------------------------------|
| Probability | 90 %    | 10 %                         |
| Payoff      | 0 coins | either 10 coins or 100 coins |

b) Censored

| Lottery     |         |          |           |
|-------------|---------|----------|-----------|
| Probability | 90 %    | 9 %      | 1 %       |
| Payoff      | 0 coins | 10 coins | 100 coins |

| Lottery draws |           |
|---------------|-----------|
| #             | Value     |
| Draw 1        | 100 Coins |
| Draw 2        | 100 Coins |
| Draw 3        | 100 Coins |
| Draw 4        | 100 Coins |
| Draw 5        | 100 Coins |

c) Sample

Fig. 2. Screenshots of the different conditions in the experiment. (a) Control condition, (b) Censored condition, (c) Sample condition.

**Willingness-to-pay.** We assume that gamers aim to maximize their subjective expected utility. Denote as  $u(z)$  the utility derived from prize  $z$ . A gamer's WTP for the loot box  $Z$  under format  $f$  is then given by

$$\mathbb{E}_{G_f}[u(Z)] = \mathbb{E}_{G^*}[u(Z)] + \underbrace{\mathbb{E}_{G_f}[u(Z)] - \mathbb{E}_{G^*}[u(Z)]}_{=: \phi(f)},$$

where  $\phi(f)$  captures a bias that operates through the gamer's subjective belief.

To link the above to our experimental design, we impose the following central assumption.

**Assumption 1.** *The control condition eliminates any bias in the WTP.*

Under [Assumption 1](#), our control condition identifies the average consumption value of a lottery. Moreover, a simple linear regression of the stated WTPs on treatment indicators identifies the average overspending on these lotteries due to censored odds and selected feedback.



**Beliefs.** To test whether any bias in WTPs operates via beliefs, we ask subjects how often they believe to win the high prize of  $x$  Coins in 100 draws. Denote as  $b_i$  the belief of subject  $i$  for a lottery  $Z$  (under format  $f$ ). We think of this belief as follows:

$$b_i = \mathbb{P}_{G_f}[Z = x] + \epsilon_i,$$

where the “noise” term  $\epsilon_i$  includes implementation errors or general optimism.

Under the assumption that this noise is independent of the format, a simple linear regression of stated beliefs on treatment indicators identifies the bias in beliefs induced by censored odds and selected feedback on the reward distribution. Note that overly “optimistic” subjects can state beliefs that contradict the objective information they have.<sup>8</sup> We will also provide analyses separately for those subjects who state realistic beliefs.

### 2.3. Implementation and logistics

As is common practice for WTP elicitation, we use the BDM mechanism to incentivize subjects (Becker et al., 1964). We do not incentivize the belief elicitation, however. Recent work by Danz et al. (2022) suggests that standard incentivization techniques (such as the binarized scoring rule) systematically distort reported beliefs. Moreover, we view the belief question as an input to (or mediator of) a subject’s stated WTP, which is our primary outcome of interest. To minimize anchoring effects, we use a slider without an initial value to elicit beliefs (see Figure A.9). To ensure that subjects engage with the lotteries, they could state beliefs (and afterwards WTPs) only after a 5 s delay.

The design was pre-registered in the AEA RCT registry as trial AEARCTR-0009501.<sup>9</sup> We collected data from 617 subjects located in the United Kingdom (UK) via Prolific in July 2022. The experiment consisted of 3 modules. First, we screened out inattentive participants via an attention check at the beginning of the experiment and after the instructions via comprehension questions. Second, all subjects who passed both tests stated their WTPs and beliefs for five lotteries. Third, we collected demographics and potential correlates of interest. Table A.6 suggest that randomization was successful. Screenshots of all parts of the experiment (including additional survey questions) can be found in Online Appendix B. Subjects earned a base fee of £1.50 for participation. In addition, 1 out of 6 participants received a bonus payment depending on the WTP stated for one randomly selected lottery. Conditional on receiving a bonus, the average bonus paid was £5.35. The experiment took, on average, 11 min to complete.

## 3. Main results

Next, we turn to our main result. Fig. 3 (left panel) displays the average WTP, separately for each treatment. First, we observe that all treatments increase the average WTP relative to the Control condition: the *Sample* condition by 43% ( $p = 0.03$ ), the *Censored* condition by 45% ( $p < 0.01$ ), and the *Joint* condition by 100% ( $p < 0.01$ ). All treatment effects are highly statistically significant and meaningful in magnitude. Common design features of loot, therefore, induce significant overspending in our context. Table 1 Column (1) confirms this result in a regression format, and Column (2) shows robustness to adding lottery fixed effects and demographic controls. Interestingly, effect sizes in *Censored* and *Sample* are statistically indistinguishable from each other ( $p = 0.78$ ), while the differences between *Censored* and *Joint* ( $p < 0.01$ ) and *Sample* and *Joint* ( $p = 0.01$ ) are statistically significant.

### Result 1. Both censoring and sampling induce overspending of almost 50%.

Furthermore, consistent with our conceptual framework, the effect of censoring and sampling operates through the belief of winning the high prize. Fig. 3 (right panel) displays the average stated belief, separately for each treatment. Relative to the Control condition, all treatments increase the average belief: the *Sample* condition by 3.1 p.p. ( $p = 0.04$ ), the *Censored* condition by 12.3 p.p. ( $p < 0.01$ ), and the *Joint* condition by 17.4 p.p. ( $p < 0.01$ ). Table 1 Columns (4) and (5) confirm these findings in regression format. Again, Column (4) displays regression results without controls, while Column (5) adds lottery fixed effects and demographic controls. Examining the pairwise differences between the three interventions we find that differential effect sizes between *Sample* and *Censored* ( $p < 0.01$ ), *Censored* and *Joint* ( $p < 0.01$ ), and *Sample* and *Joint* ( $p < 0.01$ ).

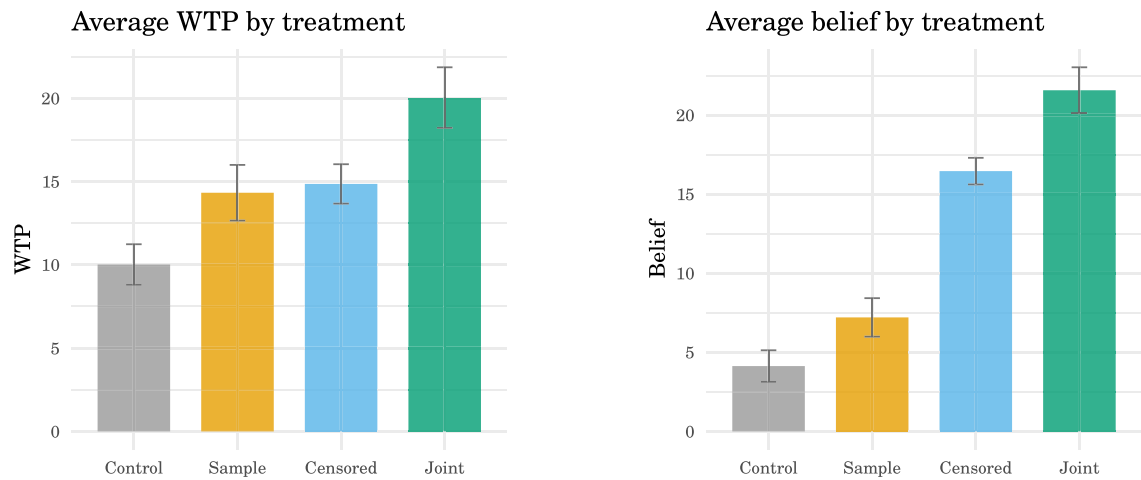
Our conceptual framework and results suggest that beliefs are the mediating channel that leads to overspending. To make this argument explicit, we control for beliefs when regressing the WTPs on the treatment indicators. Under the null of our conceptual framework, beliefs fully mediate the treatment effects on WTP. Thus, including the potential mediator in our regression allows us to test this hypothesis of full mediation. The results of this mediator analysis are in Column (3) of Table 1. Beliefs are statistically significant at any conventional level ( $p < 0.01$ ) and are positively related to the WTPs. Furthermore, the effect of *Censored*, *Sample*, and *Joint* becomes insignificant. This cleanly demonstrates that censored odds and biased samples increase the willingness-to-pay via a beliefs channel.

### Result 2. Inflated beliefs drive participants’ overspending.

One might speculate that overspending on loot boxes could be driven by self-control problems, as people with self-control problems tend to act impulsively and give in to immediate desires. The excitement and unpredictability of loot boxes can trigger this impulsive behavior, leading to excessive spending on loot boxes. Using the survey measure proposed in Tangney et al. (2004) – that tests for self-control problems at the hand of 13 established psychological questions – we find no (significant) relation between self-control problems and loot box overspending in our experiment (see Table A.2 in the Online Appendix for details).

<sup>8</sup> For example, consider the lottery that pays 0 Coins with 90% probability, 10 Coins with 9% probability, and 100 Coins with 1% probability. Any belief larger than 10% would exceed the joint probability of winning a non-zero prize and contradict the available information.

<sup>9</sup> The pre-registration can be found at <https://doi.org/10.1257/rct.9501-3.0>.



**Fig. 3.** Average willingness-to-pay and beliefs by treatment. We include all subjects that finished the experiment. WTP is the willingness to pay for a lottery. Belief is a subjective estimate of the frequency of winning the lottery in 100 independent draws. Whiskers are the standard error of the mean.

**Table 1**  
Regression results—main specification.

|              | WTP<br>No controls<br>(1) | Controls<br>(2)   | Controls + Belief<br>(3) | Belief<br>No controls<br>(4) | Controls<br>(5)     |
|--------------|---------------------------|-------------------|--------------------------|------------------------------|---------------------|
| Censored     | 4.84***<br>(1.55)         | 4.83***<br>(1.53) | 2.53<br>(1.56)           | 0.373***<br>(0.055)          | 0.373***<br>(0.055) |
| Sample       | 4.31**<br>(2.00)          | 4.17**<br>(1.99)  | 3.32*<br>(1.83)          | 0.143*<br>(0.074)            | 0.138*<br>(0.074)   |
| Joint        | 10.0***<br>(1.99)         | 10.0***<br>(2.02) | 6.04***<br>(2.05)        | 0.647***<br>(0.079)          | 0.650***<br>(0.080) |
| Belief       |                           |                   | 6.17***<br>(1.10)        |                              |                     |
| Observations | 3,085                     | 3,085             | 3,085                    | 3,085                        | 3,085               |
| Lottery FE   |                           | x                 | x                        |                              | x                   |

Notes: Results from ordinary least squares (OLS) regressions on treatment dummies. The outcome variable in columns (1), (2) and (3) is the willingness to pay and in (4) and (5) the transformed belief, i.e. the belief that the high outcome occurs conditional on the medium or high outcome being drawn. Columns (1) and (4) do not include control variables. Columns (2), (3) and (5) control for age, gender and monthly available budget. Standard errors clustered at the subject level in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## 4. Additional results

### 4.1. Correlates of overspending on loot boxes

In order to get an idea about the external validity of our findings, we also examine whether there is a relationship between our experimental measures and real-world loot-box overspending. For this purpose, we asked subjects at the end of the experiment a series of questions on their daily usage of video games and knowledge of loot boxes. Specifically, we elicited (i) whether they know what loot boxes are and if the answer is positive, (ii) how much money they spend on loot boxes per month, and (iii) whether they have ever spent more money on loot boxes than they initially planned to. In our sample, 59% play video games, documenting that gaming is a pervasive phenomenon throughout society. Our subjects spend, on average, 1.2 h playing video games daily. Moreover, 69% of our subjects know what loot boxes are. Conditional on knowing what loot boxes are, the average participant spends \$15 on loot boxes per year. This average shrouds important heterogeneity: 75% of participants do not spend at all, while the 95th percentile spender invests \$100 or more per year. A substantial minority (11%) states to have ever spent more on loot boxes than initially planned.

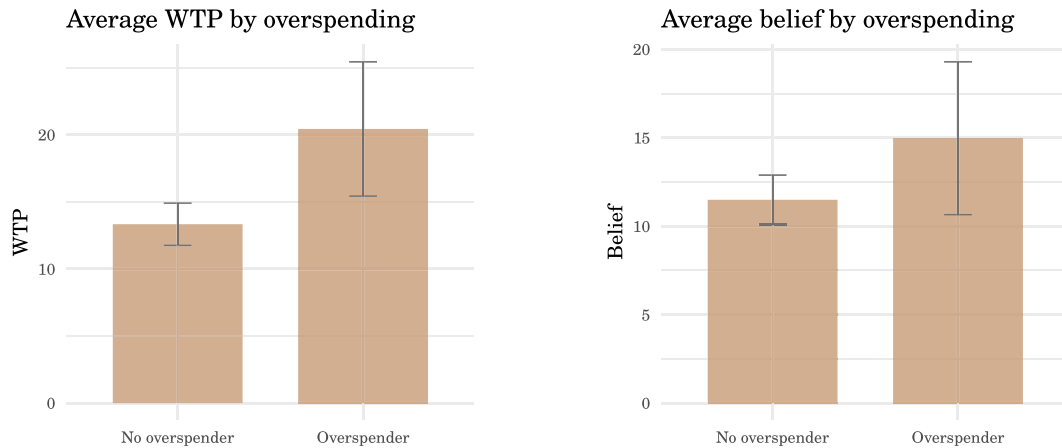
Furthermore, we elicit the gambling behavior of the subjects using the nine survey questions from the Problem Gambling Severity Index (Ferris and Wynne, 2001). Consistent with the prior literature (Zendle and Cairns, 2018), we find a positive correlation ( $\rho = 0.26$ ,  $p < 0.01$ ) between loot-box usage and survey measures of gambling behavior.<sup>10</sup> Overall, our sample thus seems well-suited to study the demand for loot boxes.

<sup>10</sup> We define subjects as loot-box users if they have either (a) ever spent more than they planned to on loot boxes or (b) have positive yearly spending on loot boxes.

**Table 2**  
Regression results—predictors for real world overspending.

|                | =1 if subject overspent on loot boxes<br>(1) | (2)              | (3)                 | (4)                 |
|----------------|----------------------------------------------|------------------|---------------------|---------------------|
| WTP            | 0.004***<br>(0.001)                          |                  |                     | 0.003**<br>(0.001)  |
| Belief         |                                              | 0.044<br>(0.033) |                     | −0.002<br>(0.034)   |
| Gambling Score |                                              |                  | 0.869***<br>(0.271) | 0.825***<br>(0.260) |
| Observations   | 425                                          | 425              | 425                 | 425                 |

Notes: Results from ordinary least squares (OLS) regressions of a dummy that is one if the subject state to have spent more than planned on loot boxes in the real world. WTP is the willingness to pay. Belief is the belief that the high outcome occurs conditional on the medium or high outcome being drawn. Gambling score is the score from a self reported gambling questionnaire, scaled from 0 to 1. All variables are subject averages. We include subjects who have stated that to know what loot boxes are. Robust standard errors in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .



**Fig. 4.** Average willingness-to-pay and beliefs for subjects that overspent on loot boxes in the past. WTP is the willingness to pay for a lottery. Belief is a subjective estimate of the frequency of winning the lottery in 100 independent draws. Whiskers are the standard error of the mean.

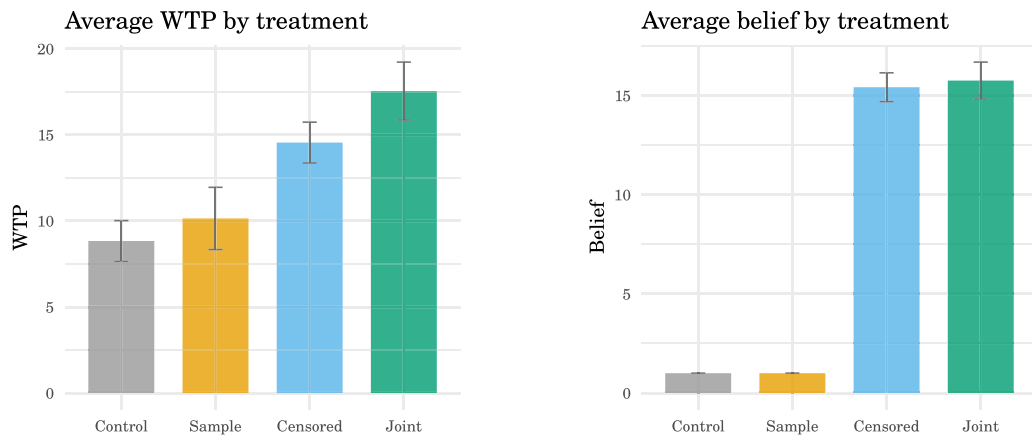
We are mainly interested in whether subjects who have overspent on loot boxes in the past differ in systematic ways from those who did not and whether our experimental measures pick up part of this variation. We asked subjects: “Have you ever spent more than you planned to on loot boxes?” We then regress a dummy variable that takes a value of one if the answer was yes and a value of zero otherwise on a subject’s average WTP and stated beliefs (see Table 2). We find a positive association between a subject’s WTP and overspending on loot boxes. An increase in the average WTP from the 5<sup>th</sup> percentile to the median (an increase of 10 coins) is associated with an increase in the probability to overspend of 3.28 percentage points. While admittedly small, the effect remains significant even when controlling for survey measures of gambling behavior. Hence, the average WTP for monetary lotteries picks up part of the variation in loot-box overspending that these survey measures cannot explain. On the other hand, stated beliefs are not correlated with the tendency to overspend on loot boxes.

Fig. 4 visualizes those results. The WTPs are higher for subjects who overspent on loot boxes in the past than those who have not: the average WTP for overspenders is 20.44, compared to a WTP for non-overspenders of 13.34 ( $p < 0.01$ ). Beliefs show a similar pattern. The average stated belief for overspenders is 14.97, while the average belief for non-overspenders is 11.50 ( $p = 0.12$ ). Note that the sample size of this analysis is comparably low ( $N = 66$  for overspenders), such that the results should be interpreted as suggestive evidence. This evidence supports the argument that overspending in the experiment reflects overspending on actual loot boxes, highlighting the external validity to the study’s findings.

#### 4.2. Treatment effects under realistic beliefs

Next, we restrict our sample to decisions in which subjects stated “realistic” beliefs that could be interpreted as conditional probabilities consistent with the information they observed. Importantly, subjects do not directly report the probability of winning the highest lottery prize. However, they provide an estimated number of wins out of 100 plays, leading to potential overestimation from being overly optimistic or perceived luck. Interpreting this belief as a winning probability leads to beliefs that appear to be unrealistic. Consider, for example, a lottery that pays 10 Coins with 39% probability and 100 Coins with 1% probability. If subjects observe this distribution (in *Control* and *Sample*), the only realistic belief is exactly 1%. In *Control*, subjects stated the realistic belief





**Fig. 5.** Average willingness-to-pay and beliefs by treatment in the main experiment. We include all subjects that finished the experiment. We exclude all decisions in which a subject stated a belief that is larger than the probability of winning a non-zero prize. WTP is the willingness to pay for a lottery. Belief is a subjective estimate of the frequency of winning the lottery in 100 independent draws. Whiskers are the standard error of the mean.

in 85% of the decisions, while in *Sample*, they stated it 62% of the time. In treatments *Censored* and *Joint*, subjects only learn the probability with which the lottery pays a non-zero prize; here, 40%. Hence, any belief between 0% and 40% is realistic in this case. This leaves us with 97% of the decisions in *Censored* and 86% of the decisions in *Joint*.

**Fig. 5** and **Table A.3** show the results. By construction of the sample, stated beliefs are identical in *Control* and *Sample*. More interestingly, conditional on stating a realistic belief, there is also no significant difference in stated beliefs across *Censored* and *Joint*. At the same time, even realistic beliefs are massively inflated (compared to the truth) in these treatments. On average, subjects assign almost equal probabilities to the events of receiving 10 Coins and  $x$  Coins. To see this, consider the right panel of **Fig. 5** and note that the highest average realistic belief across all lotteries is 30%, with a mid-point of 15%.<sup>11</sup> Next, we calculate the difference between the realistic beliefs and the mid-point of the respective lotteries for each subject. The average difference in *Censored* and *Joint* is only  $-0.44$ , which is not statistically significantly different from zero (see **Table A.4**). Furthermore, using a two-one-sided t-tests procedure for equivalence testing, we can reject the hypothesis that this difference falls outside of a five p.p. interval around zero ( $p < 0.01$  for all tests).

This is consistent with existing evidence on people – when being uncertain – biasing probabilities towards a uniform distribution (e.g., **Fischhoff and Bruine De Bruin, 1999**; **Dimmock et al., 2016**). Across the four conditions, the average WTPs follow the same qualitative patterns as before. However, neither the difference in WTP between *Control* and *Sample* nor the difference in WTP between *Censored* and *Joint* is statistically significant in this smaller subsample.

#### 4.3. Robustness experiment

One caveat of our design is that the lotteries are offered by the experimenter, not a firm trying to maximize profits. This might affect the inferences that subjects draw from observing censored probabilities or a selected sample, and it might result in higher WTPs compared to a market setting. We address this concern in a second experiment.<sup>12</sup>

In the treatment *Info*, we make it clear to subjects that the outcomes with censored probabilities are *not* equally likely, and we further provide them unbiased information on the probability distribution (on top of a selected sample). A total number of 414 subjects completed the experiment on *Prolific* in November 2022. The instructions, screenshots of the decision screens, and details on the implementation can be found in **Online Appendix B**.

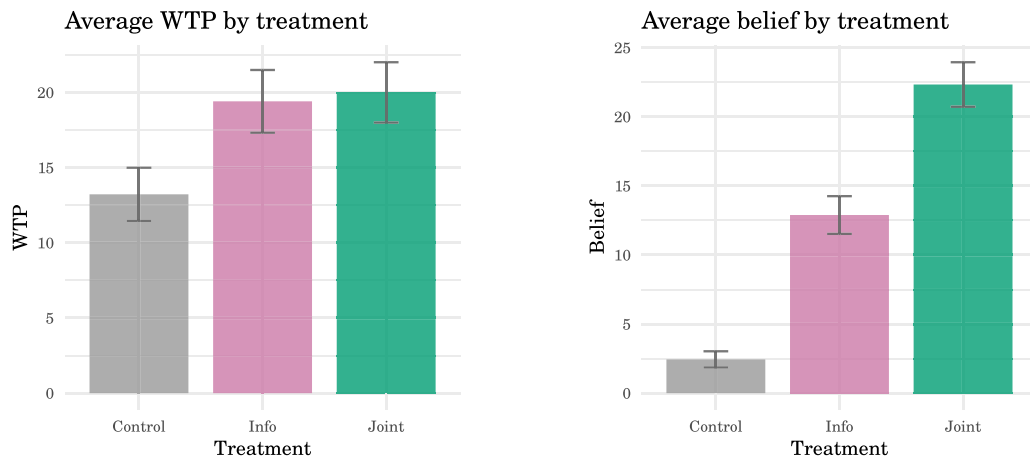
**Fig. 6** summarizes our findings. The additional information significantly decreases the average (conditional) belief of winning the high prize compared to *Joint* ( $p < 0.01$ ). While the WTP in *Info* is also slightly below the one in *Joint*, the effect is not significant at the 10% level ( $p = 0.69$ ). Importantly, both the belief and the WTP are significantly higher in *Info* compared to *Control* (see **Table A.5** in the **Online Appendix**). Overall, it highlights the robustness of our results and points to the significance of the design features of loot boxes on WTPs.

## 5. Conclusion

In a laboratory experiment, we document that two main characteristics of loot boxes, namely, censored odds and selective feedback, increase the demand and the willingness-to-pay for lotteries, arguably without providing additional utility for consumers.

<sup>11</sup> We consider lotteries that have a non-zero winning probability of  $q \in \{10, 20, 30, 40, 50\}$ .

<sup>12</sup> The pre-registration is available at <https://doi.org/10.1257/rct.10506-1.0>.



**Fig. 6.** Average willingness-to-pay and beliefs by treatment in the Robustness Experiment. We include all subjects that finished the Robustness Experiment. WTP is the willingness to pay for a lottery. Belief is a subjective estimate of the frequency of winning the lottery in 100 independent draws. Whiskers are the standard error of the mean.

We think that it is plausible to assume that we have estimated a lower bound on the distortion of loot-box demand. Real-world loot boxes are embedded into enjoyable video games and come along with contextual features that could enhance a gamer's utility. In such an environment, censored odds and selective feedback can be expected to increase gamers' demand likewise, if not even to a larger degree, which means that we estimate a lower bound on the demand-enhancing effect of loot box characteristics. For example, if gamers get distracted by fancy visual effects, they might become more prone to make statistical errors.<sup>13</sup> A nice design of loot boxes might also result in a more favorable view of the game developer and the odds it offers. Our design, which abstracts from features of loot boxes that directly generate utility, then allows us to estimate a lower bound on the bias in loot-box demand.

Our results support a case for regulating loot-box design, but it is not apparent what regulation would be effective. Current plans for regulation in Germany include labels for games with loot boxes.<sup>14</sup> However, our results show that the design of loot boxes, rather than the random rewards they provide, encourages players to overspend. Hence, this regulation may not be effective in reducing overspending. While it should be easy to enforce a transparent display of odds, it is not clear that gamers will use this information when making their purchase decisions. Our robustness experiment, for instance, suggests that additional information may not affect WTPs. Moreover, even when learning the full probability distribution over *many* prizes, gamers might not act on it because it is simply too much information to be considered. Instead, regulators must find ways of communicating the odds of loot boxes in an easily understandable way.<sup>15</sup> Preventing gamers from being confronted with selected feedback on the reward distribution comes with the additional challenge that it is not only game developers who provide it to gamers. Even if game developers are not allowed to announce prizes others have won selectively, gamers may get similar (biased) feedback from talking to their peers or watching their videos on *Youtube* or *Twitch*.

Our insights are not restricted to loot boxes but carry over to other offline and online markets. For instance, trading cards (such as Panini or Pokémon cards) and many types of online gambling use and arguably exploit the same features as loot boxes. Given our findings, similar regulations to loot boxes should be imposed on those markets to protect consumers from severe overspending. Further, our findings can also be applied to other settings, e.g., state lotteries, where the odds typically are not saliently presented to participants. However, the media often reports when a participant won the lottery – oftentimes even with articles that feature the lucky winners. Our Sample feature can be interpreted as mirroring this behavior – only information about the highest payoffs is presented to consumers. An implication of our experiment is then that news coverage about lottery winners might have harmful spillover effects on lottery consumption, as new consumers will likely overspend.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

<sup>13</sup> One dimension in which this approach and our treatments are similar is that they might increase the complexity of the economic decision problem. A recent literature examines how decision makers deal with complex environments (Enke, 2020; Oprea, 2020; Dertwinkel-Kalt and Köster, 2024).

<sup>14</sup> See, for instance, <https://usk.de/jugendschutzgesetz-aktualisiert-usk-bereitet-sich-auf-aenderungen-vor/> (accessed on September 19th, 2022).

<sup>15</sup> Alternatively, regulators could ban loot boxes altogether. As the case of Belgium shows, however, such a ban can only work if regulators also introduce proper enforcement mechanisms (Xiao, 2023).

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jebo.2024.106755>.

## Data availability

Data will be made available on request.

## References

- Barron, Kai, Huck, Steffen, Jehiel, Philippe, 2024. Everyday econometricians: selection neglect and overoptimism when learning from others. *Am. Econ. J. Microecon.* 16 (3), 162–198.
- Becker, Gordon M., Degroot, Morris H., Marschak, Jacob, 1964. Measuring utility by a single-response sequential method. *Behav. Sci.* 9 (3), 226–232.
- Benndorf, Volker, Kübler, Dorothea, Normann, Hans-Theo, 2015. Privacy concerns, voluntary disclosure of information, and unraveling: An experiment. *Eur. Econ. Rev.* 75, 43–59.
- Bolton, Patrick, Freixas, Xavier, Shapiro, Joel, 2007. Conflicts of interest, information provision, and competition in the financial services industry. *J. Financ. Econ.* 85 (2), 297–330.
- Brown, Alexander L., Camerer, Colin F., Lovallo, Dan, 2012. To review or not to review? Limited strategic thinking at the movie box office. *Am. Econ. J. Microecon.* 4 (2), 1–26.
- Chandrasekher, Madhav, Frick, Mira, Iijima, Ryota, Le Yaouanq, Yves, 2022. Dual-self representations of ambiguity preferences. *Econometrica* 90 (3), 1029–1061.
- Chen, Ningyuan, Elmachtoub, Adam N., Hamilton, Michael L., Lei, Xiao, 2021. Loot box pricing and design. *Manage. Sci.* 67 (8), 4809–4825.
- Clotfelter, Charles T., Cook, Philip J., 1990. On the economics of state lotteries. *J. Econ. Perspect.* 4 (4), 105–119.
- Danz, David, Vesterlund, Lise, Wilson, Alistair J., 2022. Belief elicitation and behavioral incentive compatibility. *Amer. Econ. Rev.* 112 (9), 2851–2883.
- Dertwinkel-Kalt, Markus, Köster, Mats, 2024. Salient cues and complexity. *Manage. Sci.*
- Deversi, Marvin, Ispano, Alessandro, Schwardmann, Peter, 2021. Spin doctors: An experiment on vague disclosure. *Eur. Econ. Rev.* 139, 103872.
- Dimmock, Stephen G., Kouwenberg, Roy, Wakker, Peter P., 2016. Ambiguity attitudes in a large representative sample. *Manage. Sci.* 62 (5), 1363–1380.
- Drummond, Aaron, Sauer, James D., 2018. Video game loot boxes are psychologically akin to gambling. *Nat. Hum. Behav.* 2 (8), 530–532.
- Drummond, Aaron, Sauer, James D., Hall, Lauren C., Zendle, David, Loudon, Malcolm R., 2020. Why loot boxes could be regulated as gambling. *Nat. Hum. Behav.* 4 (10), 986–988.
- Du, Ning, Budeşcu, David V., 2005. The effects of imprecise probabilities and outcomes in evaluating investment options. *Manage. Sci.* 51 (12), 1791–1803.
- Enke, Benjamin, 2020. What you see is all there is. *Q. J. Econ.* 135 (3), 1363–1398.
- Esponda, Ignacio, Vespa, Emanuel, 2018. Endogenous sample selection: A laboratory study. *Quant. Econ.* 9 (1), 183–216.
- Federal Trade Commission, 2020. FTC video game loot box workshop. Available at: [https://www.ftc.gov/system/files/documents/reports/staff-perspective-paper-loot-box-workshop/loot\\_box\\_workshop\\_staff\\_perspective.pdf](https://www.ftc.gov/system/files/documents/reports/staff-perspective-paper-loot-box-workshop/loot_box_workshop_staff_perspective.pdf).
- Ferris, Jacqueline Ann, Wynne, Harold James, 2001. The Canadian Problem Gambling Index. Canadian Centre on substance abuse Ottawa, ON.
- Fischhoff, Baruch, Bruine De Bruin, Wändi, 1999. Fifty-fifty=50%? *J. Behav. Decis. Mak.* 12 (2), 149–163.
- Gan, Tan, 2022. Gacha game: When prospect theory meets optimal pricing. arXiv preprint arXiv:2208.03602.
- Gilboa, Itzhak, Schmeidler, David, 1989. Maxmin expected utility with non-unique prior. *J. Math. Econom.* 18 (2), 141–153.
- Jin, Ginger Zhe, Luca, Michael, Martin, Daniel, 2021. Is no news (perceived as) bad news? An experimental investigation of information disclosure. *Am. Econ. J. Microecon.* 13 (2), 141–173.
- Jin, Ginger Zhe, Luca, Michael, Martin, Daniel, 2022. Complex disclosure. *Manage. Sci.* 68 (5), 3236–3261.
- Kachurka, Raman, Krawczyk, Michał, Rachubik, Joanna, 2021. State lottery in the lab: an experiment in external validity. *Exp. Econ.* 1–25.
- Kocher, Martin G., Lahno, Amrei Marie, Trautmann, Stefan T., 2018. Ambiguity aversion is not universal. *Eur. Econ. Rev.* 101, 268–283.
- Koehler, Jonathan J., Mercer, Molly, 2009. Selection neglect in mutual fund advertisements. *Manage. Sci.* 55 (7), 1107–1121.
- Lockwood, Benjamin B., Allcott, Hunt, Taubinsky, Dmitry, Sial, Afras, 2024. What drives demand for state-run lotteries? evidence and welfare implications. *Rev. Econ. Stud.* rdae086.
- López-Pérez, Raúl, Pintor, Ágnes, Sánchez-Mangas, Rocío, 2022. Some conditions (not) affecting selection neglect: Evidence from the lab. *J. Econ. Behav. Organ.* 195, 140–157.
- Oprea, Ryan, 2020. What makes a rule complex? *Am. Econ. Rev.* 110 (12), 3913–3951.
- Rogers, Paul, Webley, Paul, 2001. “It could be us!”: Cognitive and social psychological factors in UK national lottery play. *Appl. Psychol.* 50 (1), 181–199.
- Schmeidler, David, 1989. Subjective probability and expected utility without additivity. *Econometrica* 571–587.
- Tangney, J.P., Baumeister, R.F., Boone, A.L., 2004. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *J. Pers.* 72 (2), 271–324.
- The Norwegian Consumer Council, 2022. Insert Coin: How the Gaming Industry Exploits Consumers Using Loot Boxes. Technical Report, URL <https://storage02.forbrukerradet.no/media/2022/05/2022-05-31-insert-coin-publish.pdf>.
- Trautmann, Stefan T., Van De Kuilen, Gijs, 2015. Ambiguity attitudes. In: *The Wiley Blackwell Handbook of Judgment and Decision Making*, vol. 2, Wiley Online Library, pp. 89–116.
- Xiao, Leon Y., 2023. Breaking ban: Belgium's ineffective gambling law regulation of video game loot boxes. *Collabra: Psychology* 9 (1), 57641.
- Zendle, David, Cairns, Paul, 2018. Video game loot boxes are linked to problem gambling: Results of a large-scale survey. *PLoS One* 13 (11), e0206767.