



Mechanism Design with Moral Bidders

Shahar Dobzinski  

Weizmann Institute of Science, Rehovot, Israel

Sigal Oren   

Ben-Gurion University of the Negev, Beer-Sheva, Israel

Abstract

A rapidly growing literature on lying in behavioral economics and psychology shows that individuals often do not lie even when lying maximizes their utility. In this work, we attempt to incorporate these findings into the theory of mechanism design.

We consider players that have a preference for truth-telling and will only lie if their benefit from lying is sufficiently larger than the loss of the others. To accommodate such players, we introduce α -moral mechanisms, in which the gain of a player from misreporting his true value, comparing to truth-telling, is at most α times the loss that the others incur due to misreporting. Note that a 0-moral mechanism is a truthful mechanism.

We develop a theory of moral mechanisms in the canonical setting of single-item auctions within the “reasonable” range of α , $0 \leq \alpha \leq 1$. We identify similarities and disparities to the standard theory of truthful mechanisms. In particular, we show that the allocation function does not uniquely determine the payments and is unlikely to admit a simple characterization. In contrast, recall that monotonicity characterizes the allocation function of truthful mechanisms.

Our main technical effort is invested in determining whether the auctioneer can exploit the preference for truth-telling of the players to extract more revenue comparing to truthful mechanisms. We show that the auctioneer can indeed extract more revenue when the values of the players are correlated, even when there are only two players. However, we show that truthful mechanisms are revenue-maximizing even among moral ones when the values of the players are independently drawn from certain identical distributions (e.g., the uniform and exponential distributions).

A by-product of our proof that optimal moral mechanisms are truthful is an alternative proof to Myerson’s optimal truthful mechanism characterization in the settings that we consider. We flesh out this approach by providing an alternative proof that does not involve moral mechanisms to Myerson’s characterization of optimal truthful mechanisms to all settings in which the values are independently drawn from regular distributions (not necessarily identical).

2012 ACM Subject Classification Theory of computation \rightarrow Algorithmic game theory and mechanism design

Keywords and phrases Mechanism Design, Cognitive Biases, Revenue Maximization

Digital Object Identifier 10.4230/LIPIcs.ITCS.2022.55

Related Version *Full Version*: <https://arxiv.org/abs/2111.10674>

Funding *Shahar Dobzinski*: Work supported by BSF grant 2016192 and ISF grant 2185/19.

Sigal Oren: Work supported by BSF grant 2018206 and ISF grant 2167/19.

Acknowledgements We thank the participants of Israel Algorithmic Game Theory seminar for useful comments.

1 Introduction

The backbone of mechanism design is the assumption that individuals will lie whenever doing so increases their profit. However, a rapidly growing literature in behavioral economics and psychology questions this assumption. Many experiments demonstrate that participants often do not lie or do not lie to the full extent in order to maximize their utility. One of the



© Shahar Dobzinski and Sigal Oren;

licensed under Creative Commons License CC-BY 4.0

13th Innovations in Theoretical Computer Science Conference (ITCS 2022).

Editor: Mark Braverman; Article No. 55; pp. 55:1–55:17

Leibniz International Proceedings in Informatics



Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

now classical experiments is described in [17]: a participant is asked to roll a die in a cup and report the number to the experimenter (that cannot see which number the participant rolled). The payoff is determined according to the number reported, where higher numbers correspond to higher payoffs. Since the numbers are drawn from a known distribution, the aggregate deviation from truthful reports can be computed. A recent meta-analysis of such experiments by Abeler et al. [1] showed that participants only collect about 25% of the potential gains from lying. Similar results were observed in other experiments (e.g., reporting the number of math problems solved [29], sender-receiver games [13, 18]).

Many models were suggested to explain this *preference for truth-telling* of individuals. The meta-analysis of Abeler et al. [1] uses the data previously gathered together with new experiments to falsify many of these models. The models that survived this elimination combine a cost for lying and a desire to be perceived as honest by others.¹ The cost of lying can be interpreted as a cognitive cost [34, 6, 10]. In contrast, the desire to be perceived as honest is guided by one's preference to keep his "social identity" aligned with his moral standards, as Gneezy et al. [19] suggest. Roughly speaking, according to social identity theory (e.g., [2, 5, 33]), one's social identity is shaped by the way that others (even strangers) perceive him.

In a seminal paper, Gneezy [18] introduced the sender-receiver game to study deception. The sender-receiver game is a two-player game in which the sender recommends the receiver which alternative to choose: *A* or *B*. Each alternative implies different rewards for the sender and the receiver. The sender is asked to tell the receiver which alternative is better for the receiver. While the sender is aware of the rewards, the receiver knows only the recommendation of the sender. The rewards that each alternative offered differ over treatments. However, alternative *A* always maximized the receiver's reward, making *A* the truthful recommendation. One remarkable finding is that participants lied less in treatments where their benefit from the lie was small relative to the loss of the other player.² A possible explanation might be that the participants prefer alternative *A* over alternative *B* for other reasons. For example, Fehr and Schmidt [15] suggest that individuals have a preference for alternatives with a smaller difference in the payoff (i.e., inequality aversion). Charness and Rabin [7] suggest that individuals prefer alternatives with higher social welfare (see [16] for a comprehensive survey on such social preferences). Gneezy refutes those alternative explanations by showing that with the payoffs used in his experiments, the preference of *A* over *B* cannot be explained by the models of [15] nor [7].

Moral Mechanisms

So far, despite the large body of work demonstrating that individuals sometimes choose not to lie when it may be beneficial to them, relatively few theoretical works took this into consideration (e.g., [28, 23]). In this paper, we focus on players that rely on their moral standards when choosing their actions. In particular, such players have a preference for truth-telling (as demonstrated by [1]) and lie only if their benefit from lying is sufficiently greater than the loss of the other player(s) from lying (as demonstrated by [18, 13]).³ Due to

¹ These are also the main ingredients in two models that were published around the same time [24, 19].

² Hurkens and Karnik [21] provide an alternative explanation to the results of Gneezy [18]. They suggest that the results can be explained by assuming that some fraction of the population never lies and the rest lie whenever lying is beneficial. This criticism is refuted by Erat and Gneezy [13] in a different experimental setup.

³ The term moral is somewhat loaded, but we use it quite restrictively to refer to players that have these two preferences.

the preference for truth-telling, both profit and loss are measured with respect to truth-telling. Broadly speaking, this paper joins the growing literature that is trying to connect algorithmic game theory and behavioral economics better (e.g., [25, 27, 26, 3, 20, 14, 4, 9]).

Concretely, in an α -moral mechanism, the gain of a player from misreporting his true value, comparing to truth-telling, is at most α times the loss that the others incur due to misreporting (see Section 2 for precise definitions). Note that the case of $\alpha = 0$ coincides with the standard definition of a truthful mechanism and that the larger the value of α the more the player “cares” about the consequences of lying for other players. In this paper we limit the discussion to $0 \leq \alpha \leq 1$. Other values of α , very large and even negative, are unrealistic in many settings, but their analysis might be of mathematical interest. We leave studying these values to future work.

We stress that the behavior of moral players is very much different than the behavior of players that have externalities (such as altruistic or spiteful players, e.g., [8, 22]). The critical distinction here is that the preferences of players with externalities do not depend on their true value nor on the actions they may need to take to get the mechanism to output a preferred allocation. In contrast, moral players consider the utility of the other players only when they contemplate a lie that directly increases their own utility. As moral bidders try to follow their moral standards as much as possible, they only care about the utility of others when they are explicitly affecting it by taking a morally wrong action (e.g., lying). To give a concrete example, an altruistic player may lie to get the mechanism to output an allocation with higher social welfare even when his own utility remains the same or even decreases. In contrast, a moral player will not lie in this case.

Before diving into the technical details, it is worth discussing the applicability of moral mechanisms. Moral mechanisms are usually irrelevant in the case of competing firms. However, moral mechanisms are likely to be applicable when we have individuals that make decisions that directly influence their own payoff. In such cases, the social identity of the player is indeed shaped by his decisions. The player might therefore prefer to avoid lying, especially when the damage caused to others is large compared to the gains of the player. Naturally, this tendency is usually stronger when the players know each other personally. Thus we believe that our model is relevant when auctioning resources within an organization, or auctioning among a group of individuals that share central traits such as religion or culture, and so on. To illustrate this point, consider a hypothetical example of a department head who has to choose which PI will get some lab equipment, e.g., a fancy microscope. Each PI will naturally exaggerate the contribution of the new equipment to his lab, but we expect the exaggeration to be limited, in particular, if it is clear that another lab can make much better use of the equipment.

A Theory of Moral Mechanisms

We start with developing the foundations of moral mechanism design (Section 3). Our focus is the canonical setting of allocating a single item. We observe, as expected, that the set of allocation functions that can be implemented by moral mechanisms is a strict superset of the set of allocation functions that can be implemented by a truthful mechanism. In fact, for every $0 \leq \alpha' < \alpha \leq 1$, the set of allocation functions of α -moral mechanisms strictly contains the set of allocation functions of α' -moral mechanisms. However, in sharp contrast to dominant-strategy mechanism design, where a dominant-strategy implementable allocation function uniquely determines the payment (up to a constant), multiple payment functions might be used to implement a given allocation function as a moral mechanism.

Next, we attempt to understand the structure of morally-implementable allocation functions. A key tool in the theory of mechanism design is the characterization of the set of dominant-strategy implementable allocation functions as the set of monotone functions. We provide some evidence that an analogous characterization does not exist for moral mechanisms in the sense that there are moral mechanisms whose allocation function can be arbitrary in many “contiguous” instances.

Nevertheless, we do obtain a simple condition that characterizes the set of moral mechanisms, but this condition involves the payment function. It is well known that in a truthful mechanism, the price that the winner pays is only a function of the values of the others. That is, there are functions p_{-1}, \dots, p_{-n} and if player i wins in the instance (v_1, \dots, v_n) then his payment is $p_{-i}(v_{-i})$. We refer to mechanisms that have this property as *payment independent*. In a truthful mechanism, if player i 's profit is positive ($v_i - p_{-i}(v_{-i}) > 0$) then player i will win the item and every player $j \neq i$ is not profitable ($v_j - p_{-j}(v_{-j}) \leq 0$). In contrast, in a moral mechanism, several players might have a positive profit. This observation derives our characterization for every value of α , stated here for the simple case of $\alpha = 1$. A mechanism is a *profit maximizer* if it is payment independent and the item is allocated to a player i that maximizes $v_i - p_{-i}(v_{-i})$, as long as this expression is not negative (importantly, more than one player might have a positive profit). In particular, for $\alpha = 1$, we characterize the set of moral mechanisms as the set of profit maximizers.

The Main Result

After understanding the basic principles of moral mechanism design, we go on to our main technical effort. Suppose that the players have positive values of α . Is there a moral mechanism that generates more revenue than truthful mechanisms?⁴ In particular, can the auctioneer benefit from designing the auction in a way that will effectively “amplify” the players’ values of α ?⁵

We consider the case where the values (v_1, \dots, v_n) are drawn from a distribution \mathcal{F} in a correlated way (Section 5). We show that for *any* $0 < \alpha \leq 1$ there is a distribution \mathcal{F} such that an α -moral mechanism extracts more revenue than its optimal truthful counterpart by a multiplicative constant factor. In the other direction, we show that truthful mechanisms extract at least half the revenue of the revenue-maximizing α -moral mechanism.

In Section 4 we study the classic setting of revenue maximization with n players where the value of each player i is drawn independently from a known and identical distribution \mathcal{F} . For convenience, we assume that the valuations are discrete, i.e., there is some $\varepsilon > 0$ such that the support of \mathcal{F} is $\{0, \varepsilon, 2\varepsilon, \dots, 1 - \varepsilon, 1\}$. Our main result shows that for certain distributions, including the uniform and exponential distributions, the optimal deterministic moral mechanism is in fact truthful even when α is as large as 1. In other words, the auctioneer cannot exploit the “moral standards” of the players to increase his revenue.

⁴ In our model, bidders care about the payoffs of the other bidders but not about the payoff of the auctioneer. This is quite common as often other bidders are treated as peers whereas the auctioneer is a part of a different social group. Consider, e.g., the example discussed above of PIs who are competing for a fancy microscope. The PIs probably care about the potential benefits of their colleagues from using the microscope but not so much about “donating” money to the institutional bureaucracy.

⁵ This might be done by making the consequences of the actions of the players more explicit. For example, in a different context, Read et al. [31] asked participants to choose between taking \$5 to their pocket and donating \$5 to the red cross. When the choices were made more explicitly – “take the money and the red cross does not get the money” and “donate the money to the red cross” – the percentage of donations to the red cross increased.

Ideally, to prove our main result, we would like to extend Myerson’s celebrated characterization of optimal mechanisms [30] to hold for profit maximizers (moral mechanisms with $\alpha = 1$) and not just for truthful mechanisms.⁶ However, this approach fails almost immediately: Myerson’s first step is to characterize the payment function in terms of the allocation function. Given the characterization, Myerson’s proof completely ignores the payment functions and continues to find the allocation function that maximizes revenue. This approach inherently fails for moral mechanisms since, as we show, the allocation function of 1-moral mechanisms is “almost” unconstrained. Thus a different approach to characterizing optimal moral mechanisms is needed. Therefore, instead of “ignoring” the payment functions almost immediately and focusing on the allocation function, we develop a very different proof technique that puts the payment functions in the center. By subtle manipulations, we can characterize the payment functions of an optimal moral mechanism. We discover that the optimal moral mechanism is truthful, thus proving the optimality of Myerson’s mechanisms also among moral mechanisms.

As a by-product, we obtain an alternative proof of Myerson’s theorem for the settings that we consider. In the full version we flesh out our approach by providing an alternative proof to Myerson characterization of optimal *truthful* mechanisms when the values of the players are independently drawn from (not necessarily identical) distributions that satisfy Chung and Ely’s discrete analogue of the regularity condition [11]. Our alternative proof is not shorter than Myerson’s but is more “direct”. We – and this is obviously a subjective statement – find it easier to digest since our proof gradually “exposes” the mechanism rather than obtaining it at once via Myerson’s spectacular but somewhat opaque proof. We hope that this proof will find more applications beyond characterizing optimal moral mechanisms.

Open Questions and Future Directions

It is well documented that in practice bidders sometimes do not bid truthfully, even when truth telling is a dominant strategy. The literature underlines the importance of the specific implementation in motivating truthful bidding (e.g., ascending vs. sealed bid auctions). Our work raises the possibility of using implementations that exploit behavioral biases to encourage truthful bidding in auction settings (see footnote 5 and the discussion that leads to it). In that light, one implication of our paper is that in some settings the revenue that the auctioneer can extract does not increase with implementations that fully exploits the tendency of bidders to truth telling. It is a fascinating future direction to further explore this issue which resides on the border of behavioral economics and auction design.

We also leave several questions open on the immediate level: can moral mechanisms raise more revenue than their truthful counterparts when the distributions are not identical or not even regular? The proofs that moral mechanisms sometimes obtain more revenue and that truthful mechanisms are optimal for certain independent distributions are both quite subtle and make use of several novel technical constructs. We therefore believe that progress on this front will reveal more exciting insights. Furthermore, even considering truthful mechanisms, can the new proof technique for characterizing optimal mechanisms be used in other settings to obtain new results about truthful mechanisms, or maybe simplify existing ones?

Another promising avenue for future research is exploring moral mechanisms in other settings: can we characterize them? Are they more powerful? Even if they do not have more power, perhaps there are computationally efficient moral mechanisms for problems in which no computationally efficient truthful mechanisms exist.

⁶ Note that proving that optimal 1-moral mechanisms are truthful immediately implies that optimal α -moral mechanisms for any $0 \leq \alpha < 1$ are truthful as well.

2 Model

This paper studies single item auctions. Let $N = \{1, \dots, n\}$ be a set of players. The private information of player i is his value v_i for the item. The set of possible values of player i is \mathbb{R}^+ . An allocation function $f : (\mathbb{R}^+)^n \rightarrow N \cup \{0\}$ receives the bids of all players and decides whether to allocate the item and to which player. A direct mechanism is composed of an allocation function f and a payment function $p : (\mathbb{R}^+)^n \rightarrow \mathbb{R}^n$ that determines the payment of each player. All mechanisms in this paper are deterministic.

We study moral mechanisms for single item auctions that obey the following standard properties:

1. **Individual Rationality:** for every player i and (v_1, \dots, v_n) , $v_i \cdot \delta(i = f(v_1, \dots, v_n)) \geq p(v_1, \dots, v_n)_i$ (i.e., a player does not pay more than his value). Here, $\delta(\cdot)$ is a function that receives a boolean condition and returns 1 if the condition holds and 0 otherwise.
2. **No Positive Transfers:** For every player i and (v_1, \dots, v_n) , $p(v_1, \dots, v_n)_i \geq 0$.

Note that the two properties imply that the payment of a player that does not get the item is 0. Define the profit of player i with true value v_i for reporting value v'_i :

$$\pi_i(v'_i, v_{-i}|v_i) = v_i \cdot \delta(i = f(v'_i, v_{-i})) - p(v'_i, v_{-i})_i$$

► **Definition 1 (Moral Mechanisms).** Consider an individually rational mechanism with no positive transfers for single item auction $M = (f, p)$. M is α -moral if in every instance (v_1, \dots, v_n) , for every player i and v'_i such that $\pi_i(v'_i, v_{-i}|v_i) - \pi_i(v_i, v_{-i}|v_i) > 0$:

$$\pi_i(v'_i, v_{-i}|v_i) - \pi_i(v_i, v_{-i}|v_i) \leq \alpha \cdot \sum_{j \neq i} (\pi_j(v_i, v_{-i}|v_j) - \pi_j(v'_i, v_{-i}|v_j))$$

when $\alpha = 1$ we simply say that M is moral.

For example, when $\alpha = 1$ and all players report truthfully, if player i misreports its value then its profit might indeed increase (comparing to truth-telling), but the sum of profits of the other players decreases by at least the amount that the profit of player i increased.

Observe that any mechanism that is α -moral is also α' -moral for $\alpha' > \alpha$. In particular, note that a 0-moral mechanism is simply a truthful mechanism. To simplify the presentation, we will sometimes limit ourselves to the somewhat “cleaner” case of $\alpha = 1$. We note that our main result that shows that among all 1-moral mechanisms, the revenue-maximizing mechanism is truthful (0-moral) directly implies that for every $0 \leq \alpha \leq 1$ the revenue-maximizing mechanism among all α -moral mechanisms is truthful.

Note that truth-telling is an equilibrium of α -moral mechanisms if every player i believes that the other players report truthfully and player i will not misreport its value if its gain is less than α times the loss of the other players from this misreport⁷. In this paper we analyze the truth-telling equilibrium of α -moral mechanisms.

Our definition of moral mechanisms is inspired by [18, 19] who only consider two-player games. Two natural extensions come to mind when trying to extrapolate to n -player games: a player weighs his lie against the sum of losses of the other players, or against the maximal loss of a single player. In our single item auction setting, since the payoff of each player that does not receive the item is 0, these two extensions coincide.

⁷ This is similar to, e.g., direct revelation Bayes-Nash incentive compatible mechanisms: truth-telling is an equilibrium if all players agree on the underlying distributions from which the values are drawn and assume that the other players bid their true types.

3 Properties of α -Moral Mechanisms

We start with showing that the relaxation to α -moral mechanisms strictly extends the set of allocation functions that can be implemented:

▷ **Claim 2.** For any $\alpha > 0$ there exists an allocation function that can be implemented with an α -moral mechanism but not in dominant strategies.

Proof. Consider the following mechanism:

- If $v_1 \geq 1$, $v_1 \neq 1 + \frac{\alpha}{10}$ and $v_2 \geq \frac{1}{10}$ – allocate to player 1 at price 1.
- If $v_1 = 1 + \frac{\alpha}{10}$ – allocate to player 2 at price 0.
- Otherwise, do not allocate the item.

Note that the allocation function is not monotone, hence the mechanism cannot be implemented in dominant strategies, even with different payments. To see that the allocation function is not monotone, consider the case where $v_1 = 1$ and $v_2 = \frac{2}{10}$. The item is allocated to player 1. However, when $v_1 = 1 + \frac{\alpha}{10}$ and v_2 remains the same, the item is allocated to player 2.

We now show that the mechanism is α -moral. First, observe that truth telling is a dominant strategy for player 2: when he gets the item his payment is always zero and whether he gets the item or not is not a function of his own value. As for player 1, we first observe that if $v_2 < \frac{1}{10}$, player 1 does not get the item regardless of his report. Now, assuming that $v_2 \geq \frac{1}{10}$ we distinguish between two cases:

- When player 1 gets the item he pays 1. Thus, when $v_1 < 1$ player 1 should bid his true value. Similarly, when $1 + \frac{\alpha}{10} \neq v_1 > 1$ his profit for reporting his true value is $v_1 - 1 \geq 0$.
- If $v_1 = 1 + \frac{\alpha}{10}$, player 2 gets the item. If $v_2 \geq \frac{1}{10}$, the only way for player 1 to have a positive profit is by reporting a value that is at least 1 but not $1 + \frac{\alpha}{10}$. In this case player 1 gets the item and pays 1, so the profit of player 1 increases to $\frac{\alpha}{10}$ from 0. However, notice that player 2 gets the item and pays nothing, so his profit is at least $v_2 - 0 \geq \frac{1}{10}$. Thus the profit of player 1 indeed increases by reporting a different value, but not by more than α times the decrease of the profit of player 2. Hence the mechanism is α -moral. ◀

3.1 The Payment Functions of α -Moral Mechanisms

Truthful mechanisms have the property that the price that a player pays when it wins the item does not depend on the value that it reports. We say that a (not necessarily truthful) mechanism with this property is *payment independent*. We observe that α -moral mechanisms are payment independent.

▶ **Definition 3** (payment independence). *A mechanism $M = (f, p)$ is payment independent if the price of every winning player i does not depend on his own value. I.e., for every v_i, v'_i, v_{-i} , if player i wins in both instances (v_i, v_{-i}) and (v'_i, v_{-i}) then $p(v_i, v_{-i})_i = p(v'_i, v_{-i})_i$.*

▶ **Observation 4.** *Every α -moral mechanism is payment independent.*

Proof. Consider an α -moral mechanism $M = (f, p)$. Fix some v_{-i} and assume that there are v_i, v'_i such that the item is allocated to player i in both instances (v_i, v_{-i}) and (v'_i, v_{-i}) . Toward contradiction and without loss of generality, suppose that $p(v_i, v_{-i})_i > p(v'_i, v_{-i})_i$. Note that if player i 's value is v_i , reporting v'_i increases his profit by reducing his payment. Observe that in both instances all the other players have a zero profit since they do not win anything and hence pay nothing. Thus, the mechanism is not moral since player i can increase his profit by reporting v'_i instead of v_i without reducing the profit of any of the other players. ◀

The theory of dominant-strategy mechanisms states that in single item auctions an implementable allocation function uniquely defines the prices. In contrast, when the mechanism is 1-moral, the same allocation function may be implemented by different payment functions. For example, in the full version we show that:

▷ **Claim 5.** Consider the allocation function that allocates the item to the player with the highest value. For any $0 \leq c \leq 1$, the mechanism that allocates the item to the player with the highest value and charges him $c \cdot v_j$, where v_j is the second-highest value, is 1-moral.

Note that in the above example the higher the value of c the more revenue that the mechanism generates in every instance. In particular, among all possible choices of c , $c = 1$ gives the mechanism with the highest revenue in that family. Interestingly, when $c = 1$ the mechanism is just a second-price auction, which is obviously a dominant strategy mechanism. We show that this is not a coincidence: if f is an allocation function that can be implemented as an α -moral mechanism by several payment functions and one of the implementations is a dominant strategy implementation, then in every instance the revenue that the dominant strategy implementation generates is at least as high as the other implementations.

► **Proposition 6.** Consider a dominant-strategy mechanism $M = (f, p)$ and an α -moral mechanism $M' = (f, p')$ (both mechanisms implement the same social choice function), both are individually rational and have no positive transfers. Then, $p' \leq p$.

Proof. Recall that for any instance $v = (v_1, \dots, v_n)$ and player $i \neq f(v)$ we have that $p(v)_i = p'(v)_i = 0$. Assume towards contradiction that there exists v_i, v_{-i} such that $f(v_i, v_{-i}) = i$ but $p'(v_i, v_{-i})_i > p(v_i, v_{-i})_i$. Let v'_i be such that $p'(v_i, v_{-i})_i > v'_i > p(v_i, v_{-i})_i$. Since M is a dominant-strategy mechanism we have that f is monotone and $p(v_i, v_{-i})_i$ is the critical value of player i . This implies that player i wins the item in the instance (v'_i, v_{-i}) in both mechanisms. Since every moral mechanism is payment independent (Observation 4), in M' player i pays $p'(v'_i, v_{-i})_i = p'(v_i, v_{-i})_i$. Hence, we get that in this case $v'_i < p'(v'_i, v_{-i})_i$ in contradiction to individual rationality. ◀

3.2 Characterizations and non-Characterizations of Moral Mechanisms

Working directly with the definition of moral mechanisms in order to construct and analyze moral mechanisms is not always easy. Thus, we identify a certain natural class of moral mechanisms, called *profit maximizers*, that is easier to work with. Fortunately, we will show that α -profit maximizers are equivalent to α -moral mechanisms.

► **Definition 7 (Potential Profit).** Let $M = (f, p)$ be a payment independent mechanism. Fix v_{-i} . Let $p_{-i}(v_{-i})$ be the payment of player i if there exists some value v'_i such that $f(v'_i, v_{-i}) = i$. If such v'_i does not exist then $p_{-i}(v_{-i})$ is defined to be ∞ . $p_{-i}(v_{-i})$ is the potential payment of player i and the potential profit of player i is $v_i - p_{-i}(v_{-i})$.

► **Definition 8 (α -Profit Maximizers).** Let $M = (f, p)$ be an individually rational payment independent mechanism with no positive transfers. M is a profit maximizer if when the item is allocated, it is allocated to a player with a maximal non-negative potential profit. Furthermore, the item is allocated whenever the potential profit of at least one of the players is positive.

A mechanism is α -profit maximizer for $0 \leq \alpha \leq 1$ if in addition the following holds: for every (v_1, \dots, v_n) such that the item is allocated to player i , for every player $j \neq i$:

$$v_j - p_{-j}(v_{-j}) \leq \alpha \cdot (v_i - p_{-i}(v_{-i}))$$

We refer to 1-profit maximizers as profit maximizers. Note that, every dominant-strategy mechanism is a profit maximizer in the sense that if the potential profit of player i is positive, then his value is above his critical value (which is his potential payment). However, α -profit maximizer (for $\alpha > 0$) are more general in the sense that there might be multiple players with positive potential profits.

▷ **Claim 9.** An individually rational mechanism with no positive transfers is α -moral if and only if it is α -profit maximizing.

Proof. Consider an α -profit maximizing mechanism M . The profit of the winner is always non-negative. By no positive transfers, the profit of all other players is 0. Observe that the winner will not misreport his value since M is payment independent and hence the winner's profit can only be reduced to 0 (when he reports a value that makes him lose the item). As for the losers, let $i = f(v_1, \dots, v_n)$ and consider a loser j . Player j can misreport and change his profit from 0 to his potential profit $v_j - p_{-j}(v_{-j})$. However, in this case the winner's profit decreases by his potential profit $(v_i - p_{-i}(v_{-i}))$. Since this is an α -profit maximizer, we have that $v_j - p_{-j}(v_{-j}) \leq \alpha(v_i - p_{-i}(v_{-i}))$, as required from an α -moral mechanism. Finally, in the case that the item is not allocated (i.e., $0 = f(v_1, \dots, v_n)$) we have that the potential profit of any loser j is non-positive (i.e., $v_j - p_{-j}(v_{-j}) \leq 0$), thus player j cannot increase his profit by misreporting. Hence the mechanism is α -moral.

In the other direction, consider an individually rational α -moral mechanism with no positive transfers. Here, the only possible effect of misreporting is that the profit of the winner decreases by its potential profit and the profit of a loser increases by its potential profit. However, in an α -moral mechanism, α times the potential profit of the winner must be at least as large as the potential profit of any loser, otherwise misreporting is beneficial. This implies that the mechanism is α -profit maximizer. ◁

In the full version we rely on the equivalence of α -profit maximizers and α -moral mechanisms to prove that the payment functions that implement a given allocation function form a lattice-like structure.

A basic building block in mechanism design is the characterization of monotone allocation functions as the set of implementable functions in single parameter domains. Can we hope to obtain a similar characterization for moral mechanisms? We have already seen that when the mechanism is moral the allocation might not be monotone (Claim 2). One might hope that a different characterization of the morally implementable allocation functions exists. In the full version we show that essentially there is no simple characterization for 1-moral mechanisms. This is done by providing a family of allocation functions that are implementable using 1-moral mechanisms and their allocation can be arbitrary in a wide range of values, when $v_1 + v_2 \geq 1$ and $v_1, v_2 \leq 1$. In general, we do not know whether a characterization of the allocation functions of α -moral mechanisms exists. However, in the full version we provide a simple tool to rule out the implementability of some functions.

4 Revenue Maximization with Independent Distributions

Our main technical effort is invested in understanding whether the auctioneer can take advantage of the moral standards of the players to generate more revenue than the optimal truthful mechanism. Section 5 shows that the answer is yes when the values are drawn in a correlated manner. In contrast, we now show that moral mechanisms do not generate more revenue than truthful ones when the values are drawn i.i.d. from some set of distributions (which includes the uniform and exponential distributions).

Recall that the allocation function of moral mechanisms does not uniquely determine the payments. However, by Proposition 6, if one of the implementations of a moral mechanism is a dominant strategy one then this implementation gives the highest revenue in every instance. By itself, of course, this proposition does not guarantee that the optimal moral mechanism is truthful. For example, the mechanism given in Section 5 is moral and is not dominant strategy (and generates more revenue than any dominant strategy mechanism).

Alternatively, one could hope to extend Myerson's proof to hold for moral mechanisms as well. However, the apparent lack of structure of moral mechanisms (e.g., no condition analogous to monotonicity) is a major obstacle to implementing this approach. Instead, we develop an alternative technique that characterizes optimal moral mechanisms by directly characterizing the payment functions. Since moral mechanisms are profit maximizers, the payment functions determine the allocation function up to tie-breaking. Our proof holds for truthful mechanisms, which are 0 moral, as well. In fact, whereas for moral mechanisms we need some additional constraints on the distributions, for the simpler case of truthful mechanisms we provide in the full version an alternative proof to Myerson's characterization when values are drawn from independent (not necessarily identical) regular distributions.

For convenience, we assume that the support of the distributions is discrete. We use the following extension of virtual valuations and regularity for discrete distributions:

► **Definition 10** (Chung and Ely [11]). *Fix some $\varepsilon > 0$. Let \mathcal{F} be a distribution with support $\{0, \varepsilon, 2\varepsilon, 3\varepsilon, \dots, 1 - \varepsilon, 1\}$. Let F, f be its CDF and PDF, respectively. Let the virtual valuation of \mathcal{F} be $\varphi(v) = v - \varepsilon \cdot \frac{1 - F(v)}{f(v)}$. The distribution \mathcal{F} is regular if $\varphi(\cdot)$ is non-decreasing.⁸*

We identify a class of regular distributions for which the revenue-maximizing moral mechanism is truthful:

► **Definition 11.** *Fix some $\varepsilon > 0$. Let \mathcal{F} be a distribution with support $\{0, \varepsilon, 2\varepsilon, 3\varepsilon, \dots, 1 - \varepsilon, 1\}$. Let F, f be its CDF and PDF, respectively. We say that \mathcal{F} is standard if for every $v > v'$:*

$$v - \varepsilon \cdot \frac{(1 - F(v))}{f(v)} \geq v' - \varepsilon \cdot \frac{(1 - F(v'))}{f(v')}$$

Note the syntactic similarity of the definitions: the definition of a regular distribution is obtained by replacing the term $F(v)$ in the RHS of the definition of a standard distribution with $F(v')$. In particular, any standard distribution is also regular since for $v \geq v'$, $(1 - F(v)) \leq (1 - F(v'))$. In the full version we show that:

► **Proposition 12.** *The following distributions are standard:*

1. *The uniform distribution on $[a, b]$, for some $0 \leq a < b$.*
2. *The exponential distribution.*
3. *All distributions such that for every $v \geq v'$ in the support, $f(v) \geq \frac{f(v')}{1 + f(v')}$.*

► **Theorem 13.** *Fix $0 \leq \alpha \leq 1$. Let M be an α -moral mechanism for selling one item to n players. Suppose that the values of the players are drawn i.i.d. from a distribution \mathcal{F} and that \mathcal{F} is standard. Then, there is a dominant-strategy mechanism M' with revenue that is at least as large as the revenue of M .*

⁸ In fact, Chung and Ely assume that the support is $\{a, a + \varepsilon, a + 2\varepsilon, 3\varepsilon, \dots, b - \varepsilon, b\}$ for some $b > a$. We use $a = 0$ and $b = 1$ to slightly simplify the proof, but our proof can easily be extended to support all other values of a and b .

Since the proof is quite involved, we now provide a sketch for the 2-player case to gently introduce the reader to the main building blocks of the proof. The proof for n players is similar except for the induction hypothesis, which is more subtle in the n player case. This proof can be found in the full version.

4.1 An Outline of the Proof for 2 Players

Our goal is to prove that the optimal moral auction for two players is in fact truthful. Specifically, we will prove that this is a second-price auction with a reserve price. The reserve price equals the monopolist price of \mathcal{F} , denoted $x = \arg \max_p p \cdot (1 - F(p - \varepsilon))$ (recall that the monopolist price of \mathcal{F} is the take-it-or-leave-it price that maximizes the revenue in an auction with only one player whose value is drawn from \mathcal{F}). The main challenge is to prove the following lemma:

► **Lemma 14.** *For every player i and value $v \geq x$ of the other player we have that $p_{-i}(v) \geq v$.*

That is, if the other player's value is v for some $v \geq x$, then the potential payment of player i (i.e., the payment of player i if he wins the item) is at least v .

Before sketching the proof of Lemma 14, we apply the lemma and show that the mechanism is truthful. For the mechanism to be truthful, it is sufficient to show that for every player i and value v we have that $p_{-i}(v) \geq v$. The lemma guarantees this for $v \geq x$ and in the next paragraph we prove this for $v < x$. This is sufficient since it implies that in every instance the potential payment of the player with the lower value is at least the value of the higher value player and hence there could be at most one player with positive potential profit.

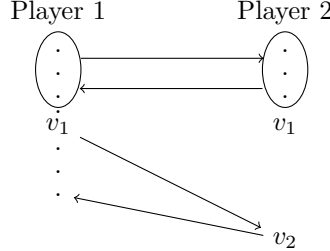
Thus, assuming Lemma 14, it remains to show that for $v < x$ we have that $p_{-i}(v) \geq v$. The intuition is as follows: by definition, the monopolist price maximizes the revenue we can extract from a player, even ignoring all other players. Denote this revenue by R_x . If the value of the other player j is $v_j < x$ and the potential payment is $p_{-i}(v_j) = x$ then player i wins the item whenever $v_i \geq x$ and pays x (because of Lemma 14 and $v_i \geq x > v_j$). Thus the revenue extracted from player i if player j 's value is $v_j < x$ is also R_x . By definition of monopolist price, there is no value of $p_{-i}(v_j)$ that extracts higher revenue.

We now sketch the proof of Lemma 14. For sake of concreteness we explicitly refer to the players here as player 1 and player 2. The proof is inductive, where the induction is over the values in the support of \mathcal{F} from the highest to the monopolist price x . Let $v_1 \geq x$ be the highest value such that for some player, without loss of generality player 2, $p_{-2}(v_1) < v_1$, i.e., when the value of player 1 is v_1 the potential payment of player 2 is less than v_1 and for any $z > v_1$, $p_{-1}(z) \geq z$ and $p_{-2}(z) \geq z$. Let $v_2 = p_{-2}(v_1)$ and consider the instance (v_1, v_2) . There are two possible cases, depending on whether $p_{-1}(v_2) \leq v_1$ or not.

In both cases we will show that we can obtain a new mechanism by increasing $p_{-2}(v_1)$ by ε , keeping the payment functions for higher values $z > v_1$ the same, and without decreasing the revenue comparing to the original mechanism. If $p_{-2}(v_1) + \varepsilon \geq v_1$, then v_1 no longer violates the lemma in the new mechanism, so we can continue and consider the next value that violates Lemma 14. Else, we analyze the new mechanism in the instance $(v_1, v_2 + \varepsilon)$ according to the two cases, and conclude that we can increase $p_{-2}(v_1)$ by another ε . We continue similarly until a mechanism with $p_{-2}(v_1) \geq v_1$ and then proceed to the next value that violates Lemma 14. At the end of the process we will have a mechanism with the property that for each player i with value $v_i \geq x$ the potential payment of the other player is at least v_i . In addition, the revenue of the new mechanism is no smaller than the revenue of the original mechanism.

Case I: $p_{-1}(v_2) \leq v_1$

This case is depicted in Figure 1. We show that in this case setting $p_{-2}(v_1) = v_2 + \varepsilon$, i.e.,



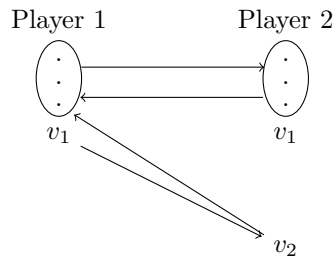
■ **Figure 1** The setting of Case I illustrated for two players. The values in the support of each player are depicted as points from low to high. A directed edge from v to u implies that the potential payment that the player with value v presents to the other player is u .

increases the potential payment by ε and does not decrease the revenue. This is shown by fixing the value of player 1 to be v_1 and comparing the expected revenue of $p_{-2}(v_1) = v_2$ and of $p_{-2}(v_1) = v_2 + \varepsilon$. Towards this end, we now analyze all instances of the form (v_1, z) and show that the revenue does not decrease in each of them:

- When $z < v_2$ player 2's potential profit is negative in both cases so the allocation and payment remain the same with both payment functions.
- When $z = v_2$ player 2's potential profit is non-positive in both cases. If player 1's potential profit is positive then player 1 gets the item and pays the same in both cases. If player 1's potential profit is 0, then when $p_{-2}(v_1) = v_2$ the potential profits of both players are 0 and player 1 will get the item by tie breaking, since his potential payment is higher. Thus, when $p_{-2}(v_1) = v_2 + \varepsilon$ player 2's potential profit becomes negative and the item is allocated to player 1 for the same price.
- When $v_1 \geq z > v_2$, we show that setting $p_{-2}(v_1) = v_2 + \varepsilon$ does not decrease the revenue. The intuition is as follows: for all values z in the range player 2's potential profit is non-negative profit even when $p_{-2}(v_1) = v_2 + \varepsilon$, so the item will be allocated. Thus, our only concern is the following instances: the item was allocated to player 2 when $p_{-2}(v_1) = v_2$ but is allocated to player 1 when $p_{-2}(v_1) = v_2 + \varepsilon$. This can happen when the potential profit of player 1 was equal to player 2's potential profit, but, due to the increase of $p_{-2}(v_1)$ the potential profit of player 1 became strictly higher. The worry is that player 1 pays less than v_2 . We claim that if player 1 becomes more profitable due to the increase in the potential payment of player 2 then the payment of player 1 is at least v_2 . This is because if the potential profits are equal then $v_1 - p_{-1}(z) = z - v_2$. However, we have that $v_1 \geq z$ thus $p_{-1}(z) \geq v_2$.
- $z > v_1$. In this case by our assumption we have that $p_{-1}(z) \geq z > v_1$. I.e., player 1 is not profitable. The potential profit of player 2 is non-negative when $p_{-2}(v_1) = v_2$ and when $p_{-2}(v_1) = v_2 + \varepsilon$ and thus player 2 will receive the item. However, the revenue is strictly higher when $p_{-2}(v_1) = v_2 + \varepsilon$ in this case.

Case II: $p_{-1}(v_2) > v_1$

This case is depicted in Figure 2. We again sketch the proof. We first show that $p_{-1}(v_2) = v_1 + \varepsilon$. The proof relies on $p_{-2}(z) \geq z$ for all $z > v_1$ and on the regularity of \mathcal{F} , and is similar to our observation above that if Lemma 14 holds then for $z < x$ we have that $p_{-1}(z) = x$.



■ **Figure 2** The setting of Case II illustrated for two players. The values in the support of each player are depicted as points from low to high. A directed edge from v to u implies that the potential payment that the player with value v presents to the other player is u .

We consider obtaining a new mechanism by applying the following simultaneous manipulations on the payment functions: $p_{-2}(v_1) = v_2 + \varepsilon$ and $p_{-1}(v_2) = v_1$. In other words, we increase $p_{-2}(v_1)$ by ε and decrease $p_{-1}(v_2)$ by ε . On one hand, the changes increase the revenue by ε in instances (v_1, z) , when $z > v_2$ (the analysis is similar to the previous case). On the other hand, the changes decrease the revenue by ε in instances (z, v_2) , $z > v_1$. When analyzing the conditions under which making both changes does not decrease the revenue the requirement that the distribution is standard organically emerges.

This concludes the sketch of proof of Lemma 14 and of Theorem 13 for the case of two players.

5 Correlated Distributions: Moral Beats Truthfulness

We now prove that moral mechanisms sometimes generate more revenue than their truthful counterparts. For this result to hold we need the values of the players to be correlated. The proof relies on the following theorem:

► **Theorem 15** (essentially [12]). *For every small enough $\delta, \varepsilon > 0$, and $1 \geq \alpha \geq \varepsilon$, there exists a distribution \mathcal{H} over the values of two players with the following properties:*

1. *The support of \mathcal{H} is finite.*
2. *All values in the support of \mathcal{H} are at least 1.*
3. *For each two possible values of player 1 v_1, v'_1 in the support of \mathcal{H} we have that $|v_1 - v'_1| > \varepsilon$.*
4. *For each possible value v_2 of player 2 in the support of \mathcal{H} we have that $1 \leq v_2 \leq 1 + \varepsilon \cdot \alpha$.*
5. *The revenue that every truthful mechanism generates on \mathcal{H} is at most $\frac{\varepsilon}{e-1} + \delta$.*
6. *There exists a revenue-maximizing truthful mechanism M in which player 2 is never allocated (i.e., all revenue is generated from player 1). In this mechanism, the item is not allocated at all with probability (over \mathcal{H}) at least $\frac{1}{e-1}$.*

We now provide some intuition for the proof of Theorem 15. Let \mathcal{H}' be a distribution that is identical to \mathcal{H} except that v_2 is always 1. The optimal mechanism for \mathcal{H}' is to make a take-it-or-leave-it offer to player 1, and if player 1 declines, sell the item to player 2 at price 1. However, in the distribution \mathcal{H} the low-order bits of v_2 determine the distribution that v_1 is drawn from. Ideally, the mechanism would query the low order bits of v_2 , determine the distribution that v_1 is drawn from, make a take-it-or-leave-it offer to player 1 and if player 1 declines sell the item to player 2 and charge 1. However, this will contradict truthfulness since player 2 can possibly misreport his low-order bits in order to fool the mechanism to believe that v_1 is drawn from another distribution, this will cause the mechanism to make a take-it-or-leave-it offer which is too high, player 1 will decline and player 2 will incur positive profit from taking the item and paying 1.

While this is only one way to fool a specific algorithm, [12] shows that every truthful mechanism can be similarly fooled. Thus, intuitively, a truthful mechanism can either (1) sell the item at a fixed price of 1 (2) make a take-it-or-leave-it offer to player 1 without querying the low-order bits of v_2 and if player 1 declines then the item is sold to player 2 at price 1, or (3) query the low order bits of v_2 , make a take-it-or-leave-it offer to player 1 and do not sell the item if he declines. The last auction is in fact the revenue-maximizing truthful mechanism.

We prove that moral mechanisms can generate more revenue on \mathcal{H} . For simplicity, consider $\alpha = 1$ and the following profit maximizer: player 1's potential payment is (almost) the best take-it-or-leave-it offer that can be made after querying player 2's low-order bits. Player 2's potential payment is 1. We show that player 1 receives the item if his profit is positive, and if his profit is negative then player 2 gets the item and pays 1. By our discussion above, this mechanism generates more revenue than the optimal truthful mechanism for \mathcal{H} . Formally:

► **Theorem 16.** *Fix $1 \geq \alpha > 0$ and $\varepsilon > 0$. There exists a distribution \mathcal{H} for which an α -moral mechanism generates a revenue of $\frac{\varepsilon}{\varepsilon-1} + \delta + \frac{1}{\varepsilon-1} - \frac{\varepsilon+1}{\varepsilon-1} \cdot \varepsilon - \varepsilon \cdot \delta$ (which is $\approx \frac{\varepsilon+1}{\varepsilon-1}$) on \mathcal{H} whereas the revenue of every truthful mechanism is at most $\frac{\varepsilon}{\varepsilon-1} + \delta$.*

Proof. Let $\mathcal{H}, \varepsilon, \delta, \alpha, M$ be as in Theorem 15. We define the following profit maximizer M' : the prices that player 2 presents to player 1 remain the same, except that we decrease them by ε . Player 1 always presents a price of 1 to player 2.

We now analyze the revenue of M' . We claim that in every instance in the support of \mathcal{H} where M allocates the item to player 1 for a payment of q , M' will also allocate the item to player 1 for a payment of $q - \varepsilon$. Moreover, when M does not allocate the item at all, M' allocates the item and charges player 2 a payment of 1. In other words, when M allocates the item to player 1 M' does so as well and the revenue barely changes. When M does not allocate the item (and recall that this happens with probability at least $\frac{1}{\varepsilon-1}$), M' generates a revenue 1. This will imply the theorem.

To prove the above claim, recall that M is truthful and thus is a profit maximizer in which at most one player has a positive profit. Let (v_1, v_2) be an instance in the support of \mathcal{H} in which player 1 is allocated and is charged q . That is, $v_1 - q \geq 0$. Consider this instance in M' . The potential profit of player 1 increases now to $v_1 - (q - \varepsilon) \geq \varepsilon$. The potential profit of player 2 is $v_2 - 1 \leq 1 + \varepsilon \cdot \alpha - 1 = \varepsilon \cdot \alpha$. We get that M' allocates the item to player 1 (if player 2 misreports his gain will be at most $\varepsilon \cdot \alpha$ and the loss of player 1 from this misreport is at least ε) and charges ε less, compared to the revenue of M . Thus, the aggregate loss of M' with respect to M is at most $\varepsilon \cdot (\frac{\varepsilon}{\varepsilon-1} + \delta)$.

Now consider an instance (v_1, v_2) in the support of \mathcal{H} where the item is not allocated at all. Note that player 2 is profitable in every instance in the support of \mathcal{H} , since $v_2 \geq 1$ and his potential payment is 1, so M' does allocate the item. If the item is allocated to player 2, then the payment is 1 and we are done. We claim that the revenue is high also when the item is allocated to player 1, despite decreasing his potential payment by ε . This follows because M is revenue maximizing, and thus the payments that player 1 pays must be in the support of \mathcal{H} (otherwise, the payments can be increased and the revenue of M increases as well). This implies that all prices presented by M are at least 1, and thus the prices presented by M' are at least $1 - \varepsilon$, as needed. The aggregate gain of M' with respect to M is at least $\frac{1}{\varepsilon-1} \cdot (1 - \varepsilon)$. Putting this together and rearranging, we get that the total gain of M' with respect to M is at least $\frac{1}{\varepsilon-1} - \varepsilon - \frac{\varepsilon+1}{\varepsilon-1} \varepsilon \cdot \delta$. ◀

Theorem 16 shows that the multiplicative gap between the revenue of an optimal truthful mechanism and an optimal moral mechanism can be as large as $\frac{e+1}{e}$. We now show that this gap cannot be larger than 2. We obtain this result by showing that the lookahead auction of Ronen [32] provides a 2 approximation not just with respect to the optimal truthful mechanism as shown in [32] but also with respect to the optimal α -moral mechanism.

► **Theorem 17.** *Fix a distribution \mathcal{H} and an α -moral mechanism M , $1 \geq \alpha \geq 0$. Denote by $rev(M)$ the expected revenue of M over \mathcal{H} . There is a truthful mechanism M' whose revenue is at least $\frac{rev(M)}{2}$.*

Proof. The mechanism M' is Ronen's lookahead auction: without loss of generality order the players so that $v_1 \geq v_2 \geq \dots \geq v_n$. Compute the marginal distribution \mathcal{H}_1 of v_1 given that $v_1 \geq v_2$. Make a revenue maximizing offer to player 1 given that his value is drawn from \mathcal{H}_1 . If player 1 accepts then he pays the take-it-or-leave-it price, otherwise the item is not sold at all. It is not hard to see (and is proved in [32]) that the lookahead mechanism is truthful.

We will now prove that the revenue of the lookahead auction is at least $\frac{rev(M)}{2}$. In fact, we may assume that M is a profit maximizer (i.e., $\alpha = 1$), this only makes our result stronger.

There are two possible cases. In the first case, at least half of the revenue that M generates is made in instances where the item is allocated to the highest value player. Here we observe that the potential profit of the highest player is non-negative. Thus, by definition, we have that the best take-it-or-leave-it offer to the highest player will generate in expectation more revenue than collecting the potential payment of player 1 (as in M), which shows that in this case, the lookahead auction provides a 2-approximation.

Else, more than half of the revenue that M generates is made in instances where the item is allocated to a player that does not have the highest value. Note that the expected revenue of M can be bounded in this case by twice the expected second-highest value. Since the take-it-or-leave-it price can be the second-highest value, the item is sold in every instance, and thus the revenue is the expectation of the second-highest value. This gives us a 2-approximation in this case as well. ◀

References

- 1 Johannes Abeler, Daniele Nosenzo, and Collin Raymond. Preferences for truth-telling. *Econometrica*, 87(4):1115–1153, 2019.
- 2 George A Akerlof and Rachel E Kranton. Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753, 2000.
- 3 Susanne Albers and Dennis Kraft. Motivating time-inconsistent agents: A computational approach. In *Proc. 12th Workshop on Internet and Network Economics*, 2016.
- 4 Moshe Babaioff, Shahar Dobzinski, and Sigal Oren. Combinatorial auctions with endowment effect. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 73–90, 2018.
- 5 Roland Bénabou and Jean Tirole. Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855, 2011.
- 6 Valerio Capraro. Does the truth come naturally? time pressure increases honesty in one-shot deception games. *Economics Letters*, 158:54–57, 2017.
- 7 Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.
- 8 Po-An Chen and David Kempe. Altruism, selfishness, and spite in traffic routing. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 140–149, 2008.

- 9 Yiling Chen, Alon Eden, and Juntao Wang. Cursed yet satisfied agents. *arXiv preprint*, 2021. [arXiv:2104.00835](https://arxiv.org/abs/2104.00835).
- 10 Shawn E Christ, David C Van Essen, Jason M Watson, Lindsay E Brubaker, and Kathleen B McDermott. The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cerebral cortex*, 19(7):1557–1566, 2009.
- 11 Kim-Sau Chung and Jeffrey C Ely. Foundations of dominant-strategy mechanisms. *The Review of Economic Studies*, 74(2):447–476, 2007.
- 12 Shahar Dobzinski and Nitzan Uziely. Revenue loss in shrinking markets. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 431–442, 2018.
- 13 Sanjiv Erat and Uri Gneezy. White lies. *Management Science*, 58(4):723–733, 2012.
- 14 Tomer Ezra, Michal Feldman, and Ophir Friedler. A general framework for endowment effects in combinatorial markets. In Péter Biró, Jason D. Hartline, Michael Ostrovsky, and Ariel D. Procaccia, editors, *EC '20: The 21st ACM Conference on Economics and Computation, Virtual Event, Hungary, July 13-17, 2020*, pages 499–500. ACM, 2020. doi: 10.1145/3391403.3399516.
- 15 Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.
- 16 Ernst Fehr and Klaus M Schmidt. The economics of fairness, reciprocity and altruism—experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1:615–691, 2006.
- 17 Urs Fischbacher and Franziska Föllmi-Heusi. Lies in disguise — an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547, 2013.
- 18 Uri Gneezy. Deception: The role of consequences. *American Economic Review*, 95(1):384–394, 2005.
- 19 Uri Gneezy, Agne Kajackaite, and Joel Sobel. Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–53, 2018.
- 20 Nick Gravin, Nicole Immorlica, Brendan Lucier, and Emmanouil Pountourakis. Procrastination with variable present bias. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pages 361–361, New York, NY, USA, 2016. ACM.
- 21 Sjaak Hurkens and Navin Kartik. Would i lie to you? on social preferences and lying aversion. *Experimental Economics*, 12(2):180–192, 2009.
- 22 Philippe Jehiel, Benny Moldovanu, and Ennio Stacchetti. How (not) to sell nuclear weapons. *The American Economic Review*, pages 814–829, 1996.
- 23 Navin Kartik, Olivier Tercieux, and Richard Holden. Simple mechanisms and preferences for honesty. *Games and Economic Behavior*, 83:284–290, 2014.
- 24 Kiryl Khalmetski and Dirk Sliwka. Disguising lies—image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics*, 11(4):79–110, 2019.
- 25 Jon Kleinberg and Sigal Oren. Time-inconsistent planning: A computational problem in behavioral economics. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pages 547–564, New York, NY, USA, 2014. ACM.
- 26 Jon Kleinberg, Sigal Oren, and Manish Raghavan. Planning problems for sophisticated agents with present bias. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pages 343–360, New York, NY, USA, 2016. ACM.
- 27 Jon Kleinberg, Sigal Oren, and Manish Raghavan. Planning with multiple biases. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, pages 567–584, New York, NY, USA, 2017. ACM.
- 28 Hitoshi Matsushima. Role of honesty in full implementation. *Journal of Economic Theory*, 139(1):353–359, 2008.
- 29 Nina Mazar and Dan Ariely. Dishonesty in everyday life and its policy implications. *Journal of public policy & Marketing*, 25(1):117–126, 2006.
- 30 Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

- 31 Daniel Read, Christopher Y Olivola, and David J Hardisty. The value of nothing: Asymmetric attention to opportunity costs drives intertemporal decision making. *Management Science*, 63(12):4277–4297, 2017.
- 32 Amir Ronen. On approximating optimal auctions. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 11–17, 2001.
- 33 Henri Tajfel. *Social identity and intergroup relations*, volume 7. Cambridge University Press, 2010.
- 34 Joseph Tao-yi Wang, Michael Spezio, and Colin F Camerer. Pinocchio’s pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, 100(3):984–1007, 2010.