Connecting the Dots: Loss Aversion, Sybil Attacks, and Welfare Maximization

Yotam Gafni* and Moshe Tennenholtz*

Abstract

A celebrated known cognitive bias of individuals is that the pain of losing is psychologically higher than the pleasure of gaining. In robust decision making under uncertainty, this approach is typically associated with the selection of safety (aka security) level strategies. We consider a refined notion, which we term loss aversion, capturing the fact that when comparing two actions an agent should not care about payoffs in situations where they lead to identical payoffs, removing trivial equivalencies. We study the properties of loss aversion, its relations to other robust notions, and illustrate its use in auctions and other settings. Moreover, while loss aversion is a classical cognitive bias on the side of decision makers, the problem of economic design is to maximize social welfare when facing self-motivated participants. In online environments, such as the Web, participants’ incentives take a novel form originating from the lack of clear agent identity—the ability to create Sybil attacks, i.e., the ability of each participant to act using multiple identities. It is well-known that Sybil attacks are a major obstacle for welfare-maximization. Our major result proves that the celebrated VCG mechanism is welfare maximizing when agents are loss-averse, even under Sybil attacks. Altogether, our work shows a successful fundamental synergy between cognitive bias/robustness under uncertainty, economic design, and agents’ strategic manipulations in online multi-agent systems.

1 Introduction

Consider an agent who needs to decide on her action in an environment consisting of other agents. In certain cases there is a uniquely defined optimal action for the agent, but in most cases this “agent perspective” is an open challenge. Given the above, both AI and economics care about an adequate modeling of an agent, and its ramifications in a variety of multi agent contexts, for example, on social welfare.

We offer a notion of loss aversion for agent modeling. As we show this has insightful effects. There are two ways to think of the solution concept we present in this paper: One is as a solution concept adapted to capture the behavioral phenomenon of the loss aversion cognitive bias in agents, particularly when probabilities over nature states are unknown. The other is as a form of robust strategy choice under uncertainty, that may be required in volatile and unpredictable environments that do not admit a stable Bayesian description. We show its usefulness in auctions, but also study its behavior in other prominent strategic settings such as facility location and voting in Appendix B. In our main result we consider the celebrated welfare maximizing VCG mechanism in combinatorial auctions setting, where it is known to fail under false name (aka Sybil) attacks. We show that loss-averse agents lead to optimal social welfare.

We use the rest of the introduction for discussion of the approach we offer, related background, as well as a careful overview of VCG under false names attack.

1.1 Cognitive Biases and Game Theory

The assumption that strategic interactions occur between rational agents that maximize expected utility is very basic to game theory as a discipline (von Neumann and Morgenstern 2007). However, this is observably wrong in many cases, and a more refined approach seeks to study games given the most realistic assumptions over agents’ behavior, where many tools that deviate from the “main line” exist to analyze such cases (e.g., trembling hand (Selten 1975), risk aversion (Cohen 1995), etc). In particular, adjustments that stem from behavioral observations are overviewed in (Rabin 2013), and include nonlinear probability weighting (Tversky and Kahneman 1992), and minimizing disappointment (Bell 1988). There is a recent interest in adapting behavioral insights from the discipline of behavioral economics into algorithmic game theory (Kleiberg and Oren 2018; Babaioff, Dobzinski, and Oren 2018; Ezra, Feldman, and Friedler 2020). Camara 2022; Kleinberg, Mullainathan, and Raghavan 2022).

Maybe the most known notion behavioral economics offers is that of loss aversion: People value gains and losses asymmetrically, so that a small loss may feel more significant than a large gain. There is prior work where loss aversion was adapted to a game theoretic framework. In (Shalev 2000), the utility functions are adapted w.r.t. reference points so that an equilibrium is formed. In (Kobberling and Wakker 2005), prospect theory is used together with concave risk aversion to adjust the utility function around a fixed reference point. We emphasize that in our framework the loss-averse strategy does not follow a notion of an equilibrium, but is individual for each player without the need to hypothesize regarding other players’ true types, choices or ratio-

*Technion - Israel Institute of Technology
nality. Most significantly, we do not assume knowledge of a distribution over types or attempt to fit within the expected utility maximization framework, but rather offer a robust notion. We compare two actions by the minimum utility of each over outcomes where they differ, and look for actions that perform well against all other actions. With regards to reference points, this can be thought of as a reference point that the agent is unwilling to compromise over. This reference point is not global across all strategies, but is a comparative measure between two fixed strategies.

Our approach is strengthened in our view by the series of works by Rabin (Rabin 2000a,b; Rabin and Thaler 2001), that point out an important conundrum of the standard modelling of risk aversion as a concave wrapper over utility functions in expectation maximization settings. It turns out that this modelling implies that agents who avoid minor risks relative to their wealth (avoiding bets where they might lose some amount \( x \) but gain \( x + \delta \)), would totally avoid any loss (no matter the possible gain) with higher stakes. Rabin sees this as a paradox that requires an alternative restatement of risk aversion outside of the expectation maximization framework. Our notion of loss aversion can be viewed as one way to do so.

### 1.2 Reasoning under Uncertainty

A classic distinction (Merrill 1982) separates reasoning under risk, where the actors are rational and there is a commonly known distribution about their environment (also known as the stochastic or Bayesian setting), and reasoning under uncertainty, where the general structure of strategies and outcomes is known, but no probabilistic information about the environment, nor even necessarily assumptions regarding actors’ rationality or behavior characteristics, are provided. For such cases, a robust or worst-case approach seems appropriate, and various notions exist to capture it. Ideally, a dominant strategy solution exists, but this is usually not the case (and indeed it is not the case in all the cases we analyze in this paper). A minimal robust notion is that of a safety level strategy, which uses a max-min approach over all possible outcomes given a strategy choice. However, though it yields interesting results in some cases (Aumann 1985; Tennenholtz 2001), in many other cases it does not tell us much about what strategy to choose, in particular in auctions settings, where we derive our most interesting results. Other refined notions are the lexicographic max-min (see, e.g., Kurokawa, Procaccia, and Shah 2018) and min-max regret (Savage 1951), which we thoroughly compare to our newly suggested loss aversion notion in Section 3 and Appendix A respectively.

### 1.3 VCG, Sybil Attacks, and Welfare

VCG is a well known mechanism which can be applied for combinatorial auctions and has good qualities such as being dominant strategy incentive compatible and achieving optimal social welfare. However, under the possibility of false-name attacks (Yokoo, Sakurai, and Matsubara 2004), it is no longer truthful. Coming up with other mechanisms does not solve the basic conundrum: In the full information settings, any false-name proof mechanism performs poorly in terms of welfare (Iwasaki et al. 2010).

A possible avenue to solving the issue is by limiting the discussed valuation classes. However, an example in (Lehmann, Lehmann, and Nisan 2006) shows that even when all bidders have sub-modular valuations, VCG is no longer dominant strategy incentive-compatible under false-name attacks. Notably though, even with this example, VCG still arrives at the socially optimal allocation, and in fact as (Alkalay-Houlihan and Vetta 2014) show this is generally true up to a constant with sub-modular (and near sub-modular) bidders. However, we show in Example 2.2 that for the XOS valuation class, which extends the sub-modular class, there is such an attack so that VCG arrives at an arbitrarily sub-optimal allocation. The attack we describe is enabled by the full information settings. Without full information, the attack is risky for the attacker, since it could lead to negative utility, as the attacker overbids her true valuation.

A useful approach, that can lead to better welfare guarantees than dominant strategy mechanism design, is Bayesian mechanism design. Assuming that the bidder distributions are common knowledge, recent work has shown that selling each item separately leads to good constant approximation welfare guarantees for XOS (Christodoulou, Kovacs, and Schapira 2016) and sub-additive (Feldman et al. 2013) valuations. Since they use false-name-proof first and second price auctions to auction items separately, their results naturally extend to Bayesian false-name mechanism design.

It is important to note, that many of the above positive results for welfare guarantees under false-name attack assume some form of risk-aversion; most importantly, that bidders do not overbid, i.e., they choose only strategies that are individually rational (under any possible nature state). This condition is equivalent to limiting the strategy space only to safety level strategies (as in this case of combinatorial auctions, the safety level is 0). In (Gafni, Lavi, and Tennenholtz 2020) the authors do not make this assumption, but their positive welfare optimality results are limited as they only consider the homogeneous single-minded with two items case. We thus believe that it is natural to ask: Under our definition of loss aversion, which is a strong risk-aversion notion (compared, e.g., to the safety level strategy), what welfare guarantees can be obtained? Surprisingly, the answer is optimal, as we show in our main result in Theorem 4.2.

### 1.4 Our Results

In Section 2, we formally define our notion of loss-aversion, and apply it to the first-price and discrete first-price auctions. In Section 3 (and the additional discussion of min-max regret in Appendix A) we describe a hierarchy of solution concepts and their relations to our loss-aversion criterion, summarized in Figure 1.

In Section 4 we present our main result. We discuss VCG as a combinatorial auction under false-name attacks, when bidders may create shill identities to send bids. It is known that VCG is not dominant strategy truthful in these settings, and previous results were limited in establishing good welfare guarantees for combinatorial auctions generally under false-name attacks. We show that when bidders...
are loss-averse, VCG achieves optimal welfare even under false-name attacks.

2 Loss Aversion in Games: A Definition and Motivating Example

Note: Missing proofs in our text appear in the full version.

When defining loss aversion, we take the perspective of a single agent \(i\) facing uncertainty. The agent has a utility function \(u_i\) that determines her utility given the state of the world, which is comprised of her own action \(a_i\), others’ actions \(a_{-i}\), and agent \(i\)’s type \(\theta_i\). Formally, \(u_i(a_i, a_{-i}|\theta_i)\). We denote by \(A_i\) the set of all agent \(i\)’s pure actions, and by \(\Delta(A_i)\) the set of all agent \(i\)’s mixed actions. An action \(a_i\) may be either from one of these action sets depending on the context. For mixed strategies, \(u_i(a_i, a_{-i}|\theta_i) = \mathbb{E}_{a_{-i}|a_i}(u_i(a_i, a_{-i}|\theta_i))\). Denote by \(\Theta_i\) the set of all agent \(i\)’s types.

**Definition 2.1.** We say that an action \(a_i\) of agent \(i\) is **loss averse** (given a type \(\theta_i\)) if for any other action \(a_i'\) over the set of outcomes where agent \(i\)’s utility differs between the actions, the minimal utility attained using \(a_i\) is at least as good as that attained by \(a_i'\). Formally, let

\[
D_{b_i}(a_i, a_i') = \{a_{-i} \text{ s.t. } u_i(a_i, a_{-i}|\theta_i) \neq u_i(a_i', a_{-i}|\theta_i)\}.
\]

Then, an action \(a_i\) is pure/mixed loss-averse if \(\forall a_i' \in A_i[1] \min_{a_{-i} \in D_{b_i}(a_i, a_i')} u_i(a_i, a_{-i}|\theta_i) \geq \min_{a_{-i} \in D_{b_i}(a_i', a_i')} u_i(a_i', a_{-i}|\theta_i)[2].
\]

We say that a strategy \(s_i : \Theta_i \rightarrow A_i\) is pure loss-averse if it maps any type \(\theta_i\) to a corresponding loss-averse pure action \(a_i\). We say that \(s_i : \Theta_i \rightarrow \Delta(A_i)\) is mixed loss-averse if it maps any type \(\theta_i\) to a corresponding loss-averse mixed action \(a_i\).

Notice that in our definition we compare pure strategies only with other pure strategies, i.e., they are loss-averse with respect to this strategy set. Mixed strategies are loss-averse w.r.t. all strategies (mixed and pure). We use the term “nature state” to mean the actions \(a_{-i}\), which may result from either uncertainty over others’ types or over their strategic choice: What matters to the agent in the end is what are all of their possible actions. There is seemingly some loss of generality in that we assume that all possible \(a_{-i}\) are fixed vectors of actions, and not more generally random variables over actions. But, as we show in Lemma C.1 allowing for this loses the usefulness of the loss-averse notion.

As a running example, we demonstrate the usage of our solution concept using the first price and discrete first price single item auctions. Interestingly, we show that in the first-price auction, there are no loss-averse strategies. However, moving to a discrete setting, we show that in the discrete first-price auction (Chwe 1989), there is a unique loss-averse strategy, which achieves maximal welfare and near maximal revenue.

**Definition 2.2.** An agent \(i\) has a value (type) \(v_i\) for an item. The agent’s bid \(b_i\) (action) and nature states \(b_{-i}\) are from the same bid space. The auctioneer allocates the item to the highest bidder (either the agent or nature, tie-breaking towards nature) and if the agent wins it receives \(v_i - b_i\), and otherwise 0.

First-price auction (FPA): bid space is \(0 \leq b_i \leq v_i\).

Discrete first-price auction (DFPA): bid space is \(b_i \in \{\epsilon \cdot k|\epsilon \cdot k \leq v_i\} k \in \mathbb{N}\).

Note that we reformulate the auctions to suit our agent perspective formulation. Moreover, we omit strategies that are not individually rational (in the (discrete) first-price auction, overbidding has negative utility in some nature states), which is justified by our later discussion in Proposition 3.4.

We also ignore multitude in nature states that does not change the auction outcome. I.e., we only consider the highest bids by others as the nature state, and not the entire bid vector. For the DFPA, we denote \(\epsilon_{net}(v_i) = \epsilon \cdot \max_{\epsilon \leq v_i} u_i\), i.e., the closest possible bid below the agent’s value of the item.

In the first-price auction, the notion of loss aversion is not of much help:

**Lemma 2.3.** In the first-price auction, there are no loss-averse bid strategies.

**Proof.** First, consider some bid \(0 \leq b_i < v_i\). Compare it with another bid \(b_i’\) that satisfies \(b_i < b_i’ < v_i\). Consider a nature state \(b_{-i}\) so that \(b_i < b_{-i} < b_i’\). Then, \(0 = u_i(b_i, b_{-i}|v_i) \neq u_i(b_i’, b_{-i}|v_i) = b_i’ - b_{-i}\). Thus,

\[
\min_{b_{-i} \in D_{b_i}(b_i, b_i')} u_i(b_i, b_{-i}|v_i) = 0.
\]

On the other hand, for the bid \(b_i’\) and for some nature state \(b_{-i}, u_i(b_i’, b_{-i}|v_i) = 0\) if and only if \(b_{-i} \geq b_i’\). In all such cases, it also holds that \(u_i(b_i, b_{-i}|v_i) = 0\). In all other cases, i.e., when \(b_i < b_{-i}\), the utility of the bidder satisfies \(u_i(b_i’, b_{-i}|v_i) = v_i - b_i’\). We conclude that

\[
\min_{b_{-i} \in D_{b_i}(b_i, b_i')} u_i(b_i', b_{-i}|v_i) = v_i - b_i' > \min_{b_{-i} \in D_{b_i}(b_i, b_i')} u_i(b_i, b_{-i}|v_i) = 0,
\]

and the bid strategy \(b_i\) is not loss-averse.

\[1\] Or, in the mixed case: \(\forall a_i' \in \Delta(A_i)\)

\[2\] We use the term minimum loosely: When taken over infinite sets that do not have a minimum the definition uses the infimum.
If \( b_i = v_i \), then for any nature state \( b_{-i} \), \( u_i(b_i, b_{-i}) = 0. \) For some \( 0 \leq b_i' < b_i \), for any nature state \( b_{-i} \) where its utility is non-zero, we have \( u_i(b_i', b_{-i}) = v_i - b_i' > 0 \), and so similarly to before \( b_i = v_i \) is not loss-averse.

However, things get more interesting with the DFPA:

**Lemma 2.4.** In the discrete first-price auction:

- For types that have \( \epsilon_{\text{net}}(v_i) \neq v_i \), bidding \( \epsilon_{\text{net}}(v_i) \) is the unique loss-averse bid.
- For types with \( \epsilon_{\text{net}}(v_i) = v_i \neq 0 \), the unique loss-averse bid is \( \epsilon_{\text{net}}(v_i) - \epsilon \).

We give a proof for the pure loss-averse case, and complete the proof for the mixed case in the full version.

**Proof.** The argument why any other bid strategy is not loss-averse follows a discretized version of the proof for Lemma 2.3.

**Case 1:** \( \epsilon_{\text{net}}(v_i) \neq v_i \)

Consider some bid with \( 0 \leq b_i' < \epsilon_{\text{net}}(v_i) \). By the same argument as in the first part of the proof of Lemma 2.3 bidding \( \epsilon_{\text{net}}(v_i) \) is loss averse w.r.t. \( b_i' \). Since there are no bids with \( \epsilon_{\text{net}}(v_i) < b_i' \leq v_i \) by the definition of \( \epsilon_{\text{net}} \), we conclude that \( \epsilon_{\text{net}}(v_i) \) is loss-averse w.r.t. all other bids, i.e., loss-averse.

**Case 2:** \( \epsilon_{\text{net}}(v_i) = v_i = 0 \)

The unique safety level bid is to bid 0, and so by Proposition 3.4 it is also the unique loss-averse strategy.

**Case 3:** \( \epsilon_{\text{net}}(v_i) = v_i \neq 0 \)

Similar to the first case, with the difference that bidding \( v_i \) always leads to utility 0, and so the loss-averse bid bracket is \( v_i - \epsilon \).

The following lemma completes the mixed loss-averse case:

**Lemma 2.5.** In the discrete first-price auction with mixed strategies, following the unique loss-averse pure strategy is the unique loss-averse strategy.

**Proof.** We show the proof for case 1 where \( \epsilon_{\text{net}}(v_i) \neq v_i \).

The other cases are done similarly.

Let \( s_i \) be the stated strategy, \( v_i \) the valuation (type) and the bid \( b = s_i(\theta_i) \). Let \( b' \) be some other bid: since \( b \neq b' \), the actualized bid \( \epsilon_{\text{net}}(v_i) \) has probability \( p < 1 \) of being the actualized bid. Consider the case \( b_{-i} \) where another bidder bids \( \epsilon_{\text{net}}(v_i) - \epsilon \), and ties are broken in favor of the other bidder. Then, \( u_i(b', b_{-i} | v_i) = \mathbb{E}_{b_{-i} | b'} [u_i(b', b_{-i} | v_i)] = p \cdot (v_i - \epsilon_{\text{net}}(v_i)) + (1 - p) \mathbb{1}[b' > \epsilon_{\text{net}}(v_i)] \cdot (v_i - b') = p \cdot (v_i - \epsilon_{\text{net}}(v_i)) + (1 - p) \mathbb{1}[b' > v_i] \cdot (v_i - b') < p \cdot (v_i - \epsilon_{\text{net}}(v_i)) < v_i - \epsilon_{\text{net}}(v_i). \)

In any nature state and actualized outcome over the mixed bid \( b' \), if \( b \) does not win the item, then \( b' \) does not win the item, or, alternatively, it wins and receives negative utility. So, \( \min_{b_{-i} \in D_{b_{-i}}(b, b')} u_i(b, b_{-i} | v_i) \geq v_i - \epsilon_{\text{net}}(v_i) > \min_{b_{-i} \in D_{b_{-i}}(b, b')} u_i(b', b_{-i} | v_i) \), and so by the loss aversion condition \( b' \) is not loss-averse (and \( b \) is loss-averse w.r.t. \( b' \)).

The simple intuition as to why the discrete first-price auction “works” (to guarantee a loss-averse strategy) and the first-price auction does not, is that in the first-price auction there is always a “safer” bid that would guarantee winning the item in more nature states. In the discrete first-price auction, due to bracketing, the highest bracket that can have positive utility is that loss-averse bid. Note that this is “almost” truthful: When \( \epsilon_{\text{net}}(v_i) \neq v_i \), it is the closest bracket to \( v_i \), and it is less than \( \epsilon \) away from it. When \( \epsilon_{\text{net}}(v_i) = v_i \) (which should be seen as a rare case, where the value precisely matches the epsilon net), it is not the truthful bracket, but it is \( \epsilon \) close to it. It is also very close to optimal revenue for the auctioneer. If \( n \) is individually rational agents participate, the most the auctioneer can get is \( \max_{1 \leq i \leq n} v_i \). If they play loss-averse strategies, they will get at least \( \max_{1 \leq i \leq n} v_i - \epsilon \).

### 3 Relations to Prominent Game-theoretic Solution Concepts

#### 3.1 Dominant Strategy

**Definition 3.1.** A weakly dominant action \( a_i \) satisfies that for any other action \( a'_i \) and any nature state \( a_{-i} \),

\[
u_i(a_i, a_{-i} | \theta_i) \geq u_i(a'_i, a_{-i} | \theta_i).
\]

A weakly dominant strategy is such that maps types to weakly dominant actions.

The following result is natural:

**Lemma 3.2.** Every weakly dominant strategy is loss-averse.

**Proof.** Fix a weakly dominant strategy \( s_i \), some type \( \theta_i \), and let \( a_i = s_i(\theta_i) \). Consider any alternative action \( a'_i \). For any \( a_{-i} \) and in particular for any such that \( u_i(a_i, a_{-i} | \theta_i) \neq u_i(a'_i, a_{-i} | \theta_i) \), it holds that \( u_i(a_i, a_{-i} | \theta_i) \geq u_i(a'_i, a_{-i} | \theta_i) \). If we assume towards contradiction that

\[
\inf_{a_{-i} \in D_{b_{-i}}(a_i, a'_i)} u_i(a_i, a_{-i} | \theta_i) = \inf_{a_{-i} \in D_{b_{-i}}(a_i, a'_i)} u_i(a'_i, a_{-i} | \theta_i),
\]

then there must be such \( a_{-i} \) with differing \( a_i, a'_i \) utilities that has \( u_i(a_i, a_{-i} | \theta_i) \leq \inf_{a_{-i} \in D_{b_{-i}}(a_i, a')} u_i(a'_i, a_{-i} | \theta_i) \), in contradiction to the dominance condition.

#### 3.2 Safety Level Strategy and Individual Rationality

Safety level strategies in non-cooperative games are such strategies that yield a best possible guarantee of utility for a player, without the need to reason about the types or strategies chosen by other players. An example by Aumann [1985] makes a compelling argument for choosing such strategies: There are games where the Nash Equilibrium does not guarantee more than the safety level. In such cases, choosing the equilibrium strategy runs the unnecessary risk of a lower outcome. Tennenholtz [2002] extends this insight and shows a class of games where the safety level strategy guarantees a large constant fraction of the Nash equilibrium outcome, without its involved risks.

Individual rationality is a common requirement in game theory analysis (see, e.g., [Nisan et al. 2007]), that requires either that an agent does not participate in a game where it
gains negative utility, or that it does not choose a strategy that may yield negative outcomes. We define:

**Definition 3.3.** A safety level strategy \( s_i \) is a strategy (mixed or pure) of player \( i \) such that for any type \( \theta_i \), it chooses an action \( a_i \) so that for any nature state \( a_{-i} \) of the other agents, 
\[
 u_i(a_i, a_{-i} | \theta_i) \geq \max_{a_i} \min_{a_{-i}} u_i(a, a_{-i} | \theta_i).
\]

\( i.e. \), the strategy guarantees the safety level \( L \) \( \text{def} \) \( \max_{a_i} \min_{a_{-i}} u_i(a', a_{-i} | \theta_i) \).

**Proposition 3.4.** A loss-averse strategy is a safety level strategy, but not necessarily vice-versa.

**Proof.** (Loss aversion \( \implies \) safety level)

Assume towards contradiction a strategy \( s \) is loss-averse but not safety level. There is always guaranteed to exist a safety level strategy \( s' \). We show that \( s \) is not loss-averse w.r.t. \( s' \). Since \( s \) is not safety level, there is a type \( \theta_i \) and nature state \( a_{-i} \) where the action \( a = s(\theta_i) \) satisfies \( u(a, a_{-i}) < L \), where \( L \) is the safety level. For that setting (as in every setting), with \( a' = s'(\theta_i), u(a', a_{-i}) \geq L \). Therefore,
\[
\min_{a_{-i} \in D(a, a')} u_i(a, a_{-i}) < L \leq \min_{a_{-i} \in D(a, a')} u_i(a', a_{-i}),
\]

where the last inequality is again by the safety level property of \( s' \).

(Safety level \( \not\implies \) loss aversion)

In the first-price auction, there are infinite safety level strategies (namely, every bid with \( 0 \leq b_i \leq v_i \)), but Lemma 2.3 shows that there are no loss-averse strategies. Similarly, in the discrete first-price auction, there may be a finite (but larger than 1) amount of safety level strategies, but a unique loss-averse strategy, by Lemma 2.4.

**Corollary 3.5.** When there is a finite amount of safety level strategies, and a finite amount of nature states, a loss-averse strategy is guaranteed to exist.

The corollary is a result of Lemma 3.9 and Lemma 3.10. We prove both during our discussion of the lexicographic max-min in the next subsection.

### 3.3 Lexicographic Max-min

A very interesting comparison is with another robust solution notion, the lexicographic max-min (also commonly known as leximin). The leximin is especially prevalent in the fair allocation literature, see, e.g., *(Kurokawa, Procaccia, and Shah 2018)*. We consider two possible ways to define it:

**Definition 3.6.** Leximin - Let \( U_{a_i} \) be the set of all possible utility outcomes of the action \( a_i \) by agent \( i \), ordered from small to large, and let \( U_{a_i}[j] \) be the \( j \) element of \( U_{a_i} \) in this ordering. An action \( a_i \) lexicographically weakly dominates (LD) another action \( a'_i \) if \( \min U_{a_i} > \min U_{a'_i} \), or \( \min U_{a_i} = \min U_{a'_i} \) and \( U_{a_i} \setminus U_{a'_i} \text{ LDs } U_{a'_i} \setminus U_{a_i} \) (a recursive definition). We call an action that LDs all other actions a leximin. A strategy is leximin if it maps all types to leximin actions.

**Multi-Leximin** - Let \( U_{a_i} \) be the multiset of all possible utility outcomes of the action \( a_i \) by agent \( i \), ordered from small to large. The rest of the definition follows similarly, where importantly in the recursive definition we remove only one copy of the minimum element at each step.

Note that the (Multi-)leximin notions are only clearly defined when there is a finite amount of nature states \( a_{-i} \), otherwise the recursive definition of LD may not terminate.

We first note that both definitions give stronger notions than safety level strategies.

**Lemma 3.7.** (Multi-)leximin is a safety level strategy, but not necessarily vice-versa.

**Proof.** (Leximin \( \implies \) safety level)

Let \( s_i \) be a leximin strategy, \( \theta_i \) some type, and \( a_i = s_i(\theta_i) \). Whether the leximin is with or without multiplicities, we have that \( \min_{a_{-i}} u_i(a_i, a_{-i} | \theta_i) = \min_{a_{-i}} u_i(a_i, a_{-i} | \theta_i) \min_{a_{-i}} u_i(a_i, a_{-i} | \theta_i) \) for any alternative action \( a'_i \), which implies that \( \min_{a_{-i}} u_i(a_i, a_{-i} | \theta_i) \geq \max_{a_{-i}} \min_{a_{-i}} u_i(a'_i, a_{-i} | \theta_i) \), which is the condition for a safety level strategy.

(Safety level \( \not\implies \) leximin)

Consider the normal-form game with two players and strategies \( a, b \) (for both). If \( u_1(a, b) = u_1(a, a) = v_i, u_1(b, b) = 10 \), then both strategies \( a, b \) are safety level for player 1 but only the strategy \( b \) is leximin (both with and without multiplicities).

Despite some similarity in the definition with loss aversion, the notion of leximin does not have a special relationship with it: neither implies the other.

**Example 3.8.** We demonstrate that leximin is different from loss aversion using the discrete first-price auction. Bidding 0 is the leximin action, as its set of outcomes is simply the set of two items \( U_0 = \{0, v_i\} \). This is the leximin since any other bid \( b_i > 0 \) has \( U_{b_i} = \{0, v_i - b_i\} \).

The notion of multi-leximin is much more closely related to the loss aversion notion. In fact, we can show that it is a stronger notion:

**Lemma 3.9.** Multi-leximin is a loss-averse strategy, but not necessarily vice-versa.

**Proof.** (Multi-leximin \( \implies \) loss-averse)

Let \( s_i \) be a multi-leximin strategy, \( \theta_i \) a type, and let \( a_i = s_i(\theta_i) \). Assume towards contradiction that \( a_i \) is not
loss-averse and there is some action $a'_i$ with
\[
m = \min_{a_i \in D_{a_i}(a_i, a')_i} u_i(a_i, a_{-i}|\theta_i) < \min_{a_i \in D_{a_i}(a_i, a'_i)} u_i(a'_i, a_{-i}|\theta_i) \overset{\text{def}}{=} m'.
\]

Let $C$ be the number of nature states where $u_i(a_i, a_{-i}|\theta_i) = u_i(a'_i, a_{-i}|\theta_i) = m$. Let $j$ be the index of the $C + 1$ appearance of $m$ in $U_{a_i}$. For any index $1 \leq j' < j$, it must be either that $U_{a_i}[j'] < m$ or that it is one of the $C$ nature states where $u_i(a_i, a_{-i}|\theta_i) = u_i(a'_i, a_{-i}|\theta_i) = m$ and so it must be an outcome with $u_i(a_i, a_{-i}|\theta_i) = u_i(a'_i, a_{-i}|\theta_i)$. Since $n' > m$, we thus know that $U_{a'_i}$ is the same as $U_{a_i}$ up to (and including) index $j - 1$. At index $j$, it must have a different outcome than under $a_i$, by our choice of $j$ using $C$. Thus, its outcome must be $m'$ (the lowest outcome that is different than under $a_i$), which means it lexicographically dominates $a_i$, in contradiction to $a_i$ being loss-averse.

\begin{mdframed}
\textbf{Lemma 3.10.} When there is a finite amount of safety level strategies, a multi-leximin is guaranteed to exist.
\end{mdframed}

\textit{Proof.} Fix a type $\theta_i$. By Lemma 3.7 any safety level action LDs all non safety level actions. Thus, if we find an action that LDs all safety level actions, this action is the leximin.

For any safety level action $a_i = s_i(\theta_i)$, consider the multi-set $U_{a_i}$. Now, lexicographic domination induces a full order over the safety level actions and there must be at least one maximizing strategy, which is the leximin.

We recall that we used the above two lemmas to derive Corollary 3.5. For the discrete first-price auction, this yields an immediate corollary:

\begin{mdframed}
\textbf{Corollary 3.11.} The unique loss-averse strategy of the discrete first-price auction is also the unique multi-leximin strategy.
\end{mdframed}

\textit{Proof.} For an agent $i$ with value $v_i$, there is a finite amount of safety level strategies, namely all the strategies with $b_i \leq v_i$, the amount of which is at most $\left\lceil \frac{v_i}{\epsilon} \right\rceil + 1$. By Lemma 3.10 there must exist a multi-leximin strategy. By Lemma 3.5, it is also loss-averse. Since there is a unique loss-averse strategy by Lemma 2.4, it must also be the unique multi-leximin.

An important advantage of the loss aversion definition is that it naturally extends to settings with continuous outcomes. It is not clear how to extend the leximin definition to such cases. Thus, one possible way of thinking about the loss aversion notion is that it is a somewhat weaker notion of multi-leximin, that can be used in continuous settings, as well as discrete ones.

To complete the picture, we note the relation of (multi-)leximin to dominant strategy. It is straightforward to show that dominant strategy implies multi-leximin. Interestingly, this does not hold for leximin:

\begin{example}
\textbf{Dominant strategy }\Rightarrow \textbf{ leximin}
\end{example}

Consider two players, one with actions $a$, $b$, the other with actions $A, B, C$. Consider $u_1(a, A) = 0 = u_1(b, A) = u_1(b, B), u_1(a, B) = 1, u_1(a, C) = 5, u_1(b, C) = 3$. Action $a$ strongly dominates action $b$ for player 1. However, the outcomes without multiplicity sets are $S_a = \{0, 1, 5\}, S_b = \{0, 3\}$, and so $b$ is the unique leximin.

\section{Main Result: Application to VCG under False-name Attacks}

False-name attacks by an agent $i$ in a combinatorial auction are where instead of sending one combinatorial bid, the agent sends multiple combinatorial bids (a vector $b_i$ rather than a single bid $b_i$). The agent then gets all the items allocated to the “agents” (which we call Sybil agents or Sybil bids) $1 \leq j \leq |b_i|$, and pays the sum of all their payments. Before formally introducing the VCG notations, we note three complexities that are present in our notations: (1) We consider both loss-aversion (which has the single agent perspective vs. nature states) and social welfare (which accounts for $n$ different agents). (2) We consider welfare for the real $n$ underlying agents of the auction, but since each may use Sybil identities, the VCG allocations are in terms of the Sybil identities. We allow for both by using sub-indexing. (3) Similar to the case of the first-price auction, discretization of the bid space is essential to the result (a counter-example for continuous VCG is in Example 3.3). To further simplify the proof, we also assume that the valuation space is discrete, though this assumption can be removed.

We allow more granularity to the bid space: valuations are on an $\epsilon$ grid, while bids are on an $\epsilon/2$ grid.

\begin{definition}
Grid(\epsilon) = \{\epsilon k \mid k \in \mathbb{N} \} = \{0, \epsilon, 2\epsilon, \ldots\}. A combinatorial bid $b \in B$ over an item set $M$ is a function $b : P(M) \rightarrow \text{Grid}(\frac{\epsilon}{2^{|M|}})$ from the power set of all subsets of $M$ to a non-negative bid value. A combinatorial valuation $v$ is similarly $v : P(M) \rightarrow \text{Grid}(\epsilon)$. With the possibility of Sybil attacks, an agent $i$ with valuation (type) $v_i$ sends a vector of bids (action) $b_i \in B^*$ (i.e., any amount of combinatorial bids), and faces a nature state $b_{-i} \in B^*$.

Let $\eta_i = |b_i|, \eta_{-i} = |b_{-i}|$ be the number of (Sybil) agents in each vector: An allocation $\alpha^S(b_i, b_{-i})$ maps the bid vectors to a partition of $S$ into subsets. We allow indexing $\alpha_{i_1, \ldots, i_n}$ to mean the union of items allocated to the Sybil identities of each real agent, as well as sub-indexing $\alpha_{i_1, \ldots, i_n}$ to mean the items allocated to a specific Sybil identity of agent $i$. We denote $\text{SW}_{\text{Obs}} = \sum_{i=1}^n \sum_{j=1}^{\eta_i} b_{ij}(\alpha_j), \text{SW}_{\text{Real}} = \sum_{i=1}^n v_i(\alpha_i)$ for the observed social welfare of an allocation as specified in the (possibly Sybil) bids, and the real social welfare of the agents, respectively. We denote $\text{truth}_i = v_i$ for the truthful bid.


The VCG combinatorial auction is the pair of allocation rule
\[ \alpha^M(b_1, b_{-1}) = \arg \max_{a^M(b_1, b_{-1})} (SW_{Ob\alpha^M(a^M, b_{-1})}), \]
and the payment rule
\[ p_{ij}^M(b_1, b_{-1}) = SW_{Ob\alpha^M(a^M, b_{-1})} - SW_{Ob\alpha^M(a^M', b_{-1})}. \]
Finally, the utility of agent \( i \) is \( u_i(b_1, b_{-1}) = v_i(S) \).

**Theorem 4.2.** When all bidders play loss-averse strategies, discrete VCG achieves optimal welfare, even under the possibility of false-name attacks and with general valuations.

**Proof.** Our proof follows the following structure: First, we define underbidding Sybil attacks and show that they are not loss-averse. We then define underbidding attacks and show that they are not loss-averse. For any of the remaining attacks, which we call exact-bidding (bidding truthfully is also exact-bidding), we show that even though they are not necessarily truthful, they yield maximal welfare. However, this still does not guarantee that one of the remaining strategies is in fact loss-averse. For this purpose, we show that there exists a loss-averse strategy: being truthful.

First, we show that if the Sybil bids are underbidding \( v \) (in a sense that will be immediately defined), then, similarly to our proof for the first-price auction, it is not safety level and thus not loss-averse. This requires slightly more care since the bids are combinatorial and there are several Sybil bids. We say that \( b_t = (b_{t1}, \ldots, b_{tn}) \) is underbidding if there is a set \( S \) and an allocation \( \alpha^S(b_t) \) so that \( \sum_{j=1}^n b_{ti}(\alpha^S_{ij}(b_t)) > v_i(S) \).

**Claim 4.3.** Underbidding \( \Rightarrow \) not loss-averse.

**Proof.** Let \( b = \max_{1 \leq j \leq n} \max_{S \subseteq M} b_{ij}(S) + v_j(S) \) for a number high enough that if some other agent bids it for any subset of \( M \), both the truthful bid \( v_i \) or the Sybil attack \( b_t \) will lose that subset. We will use it in our construction of nature states. By the underbidding condition, we can take the average \( b = \frac{v_j(S)}{2} + \frac{1}{2} \sum_{j=1}^n b_{ij}(\alpha^S_{ij}) \), so that \( \sum_{j=1}^n b_{ij}(\alpha^S_{ij}) > b > v_i(S) \). Consider a nature state where the false-name attacker faces exactly one additive bidder \( b' \) that has for any good \( g \in M \setminus S \), \( b'(g) = \frac{b}{|S|} \), and for any good \( g \in S \), \( b'(g) = \frac{b}{|S|} \). The optimal observed welfare allocation is to allocate all goods in \( M \setminus S \) to \( b' \), and allocate the set \( S \) as in \( \alpha_1(S) \). The payment of bidder \( b_t \) must be at least \( b'(S) = \frac{b}{|S|} > v_i(S) \). Therefore, the attacker has negative utility in this case, while truthfulness is individually rational: i.e., it is not a safety level strategy and so also not loss-averse.

We say that \( b_{ti}, b_{i\alpha_j} \) are underbidding if there is a set \( S \) so that for any allocation \( \alpha^S \) defined \( \alpha^S(b_t) \) so that \( \sum_{j=1}^n b_{ij}(\alpha^S_{ij}) < v_i(S) \).

**Claim 4.4.** Underbidding \( \Rightarrow \) not loss-averse.

**Proof.** Let \( b = \frac{1}{2} \sum_{j=1}^n b_{ij}(\alpha^S_{ij}) + \frac{v_j(S)}{2} \), then
\[ \sum_{j=1}^n b_{ij}(\alpha^S_{ij}(b_t)) < b < v_i(S). \]

Let \( b' \) be constructed as in the overbidding case. The allocation \( \alpha^M(b_1, b') \) allocates no items to the Sybil bidders of agent \( i \). However, the allocation given agent \( i \) bids truthfully \( \alpha^M(\text{truth}_i, b') \), allocates the set \( S \) to her with payment \( b \), which yields agent \( i \) a positive utility \( v_i(S) - b \). This yields
\[ \min_{b_{-1} \in D_{\alpha}(b_1, \text{truth}_i)} u_i(b_1, b_{-1}) = 0. \]

On the other hand, we claim that since we know loss-averse strategies are not overbidding, there are no nature states for which an underbidding Sybil attack gets positive utility while bidding truthfully gets 0 utility. Assume towards contradiction \( \text{truth}_i \) gets 0 utility. It then either does not win any item, or wins some set \( S \) and pays \( v_i(S) \) for it. Let \( S \), be the set that the Sybil bidders win to gain positive utility. As there is no overbidding, this set can be won by \( \text{truth}_i \) as well (in the respective maximizing allocation). Then,
\[ SW_{Ob\alpha^M}(b_1, b_{-1}) \]
\[ = SW_{Ob\alpha^M}(\text{truth}_i, b_{-1}) \]
\[ = SW_{Ob\alpha^M}(b_{-1}) \]
\[ = SW_{Ob\alpha^M}(b_{-1}) - v_i(S) \]
\[ \leq SW_{Ob\alpha^M}(b_1, b_{-1}) - v_i(S) \]
So
\[ v_i(S) \leq SW_{Ob\alpha^M}(b_1, b_{-1}) - SW_{Ob\alpha^M}(b_{-1}) \]
(1)

Since our choice of \( S \) assumes the Sybil bids win exactly it, we have
\[ SW_{Ob\alpha^M}(b_1, b_{-1}) + SW_{Ob\alpha^M}(b_{-1}) = SW_{Ob\alpha^M}(b_{-1}), \]
and so, together with Eq. [1]
\[ v_i(S) \leq SW_{Ob\alpha^M}(b_1, b_{-1}) - SW_{Ob\alpha^M}(b_{-1}) = SW_{Ob\alpha^M}(b_{-1}). \]

Since there is no overbidding, \( SW_{Ob\alpha^M}(b_{-1}) = v_i(S) \).

We now show that any Sybil bidder \( j \) pays \( v_i(S) - \sum_{1 \leq l \neq j} b_{ji}(\alpha^S_{ij}) \). Since \( S \) is allocated to the Sybil bidders and \( M \setminus S \) to others,
\[ SW_{Ob\alpha^M}(b_1, b_{-1}) = SW_{Ob\alpha^M}(b_1, b_{-1}) + SW_{Ob\alpha^M}(b_1, b_{-1}) \]
(2)

\[ \text{Another, albeit non-constructive method to show there exists a loss-averse strategy is by showing the finiteness of undominated exact-bidding Sybil attacks, and then use Corollary 3.3.} \]
\[ p_{ij}^M = SW^{\text{Obs}}_{\alpha_M} (b_i, b_{-i}) - SW^{\text{Obs}}_{\alpha_M b_j} (b_i, b_{-i}) \]
\[ = SW^{\text{Obs}}_{\alpha_M} (b_i, b_{-i}) - SW^{\text{Obs}}_{\alpha_M b_j} (b_i, b_{-i}) - SW^{\text{Obs}}_{\alpha_M b_j} (b_i, b_{-i}) \]
\[ = v_i (S) - SW^{\text{Obs}}_{\alpha_M b_j} (b_i, b_{-i}) \]
\[ = v_i (S) - \sum_{j=1}^{\eta_i} b_j (\alpha_{ij}) \]

The total payment of agent \( i \) is then
\[ \sum_{j=1}^{\eta_i} p_{ij}^M = \sum_{j=1}^{\eta_i} v_j (S) - \sum_{j=1}^{\eta_i} b_j (\alpha_{ij}) = \eta_i \cdot v_i (S) - (\eta_i - 1) \sum_{j=1}^{\eta_i} b_j (\alpha_{ij}) = v_i (S). \]

This concludes that whenever the utility of \( truth_i \) is 0, then the utility for the Sybil attack is 0 as well.

In any other case, the utility of \( truth_i \) must be strictly positive, and since the bids are discrete the minimum over all these cases satisfies
\[ \min_{b_{-i} \in D_{v_i} (b_i, truth_i)} u_i (truth_i, b_{-i} | v_i) \geq \frac{1}{2 |M|}. \]

Therefore, underbidding is not loss-averse. \( \square \)

We consider exact-bidding such Sybil bids that have for any set of items \( S \),
\[ \max_{a_j b_j} \sum_{j=1}^{\eta_i} b_j (\alpha_{ij} (S)) = v_i (S). \]
These are exactly all the Sybil attacks that are neither overbidding nor underbidding. \( truth_i \) is also exact-bidding.

**Claim 4.5.** Exact-bidding \( \implies \) optimal welfare.

**Proof.** Consider an allocation \( \alpha_F \overset{\text{def}}{=} a_M (b_i, b_{-i}) \) attained by all players choosing an exact-bidding attack, vs \( \alpha_T \overset{\text{def}}{=} a_M (truth_i, truth_{-i}) \). We have
\[ SW^{\text{Real}}_{\alpha_F} \leq \text{ (Truthful) } \]
\[ SW^{\text{Obs}}_{\alpha_F} \leq \text{ (No underbidding) } \]
\[ SW^{\text{Obs}}_{\alpha_T} \leq \text{ (No overbidding) } \]
\[ SW_{\alpha_T} \leq \text{ (Truthful) } \]

In words, since there is no underbidding in the Sybil attack, if we take the set allocated to each agent \( i \) under the allocation that maximizes welfare under truthfulness, there are Sybil bidders \( i_1, \ldots, i_k \) with the same aggregate valuation for it. So, \( SW^{\text{Obs}}_{\alpha_T} \) is lower bounded by the optimal truthful welfare. Since there is also no overbidding, whatever allocation is chosen as \( \alpha_F^{\text{Obs}} \) is at least as good to each agent \( i \) as is declared. \( \square \)

**Claim 4.6.** \( truth_i \) is loss-averse.

**Proof.** Consider some exact-bidding Sybil attack \( b_j \).

Case 1: There is a set \( S \) so that \( \forall 1 \leq j \leq \eta_i, b_j (S) < v_i (S) \). Then, by the exact-bidding condition there must be some allocation \( \alpha^{\mathcal{S}}_F (b_j) \) (with at least two non-empty allocations \( \alpha^{\mathcal{S}}_F \)) so that \( \max_{j=1}^{\eta_i} b_j (\alpha^{\mathcal{S}}_F (b_j)) < \sum_{j=1}^{\eta_i} b_j (\alpha^{\mathcal{S}}_F (b_j)) = v_i (S) \). Consider the nature state where there is one bid \( b' \) so that \( b' (\alpha^{\mathcal{S}}_F (b_j)) = v_i (S) \) for any \( 1 \leq j \leq \eta_i \), and the rest of the sets are defined upward-monotonically: They inherit the largest value of a subset. With this nature state, the Sybil attack has utility 0. On the other hand, \( truth_i \) has positive utility of \( v_i (S) - \max_{j=1}^{\eta_i} v_i (\alpha_{ij}) > 0 \). Since \( truth_i \) is individually rational, it is thus loss-averse w.r.t. such Sybil attacks.

Case 2: For every set \( S \), there is such \( b_j \) with \( b_j (S) = v_i (S) \). It must hold by the exact-bidding condition that for any allocation \( \alpha^{\mathcal{S}}_F (b_j) \), \( \sum_{j=1}^{\eta_i} b_j (\alpha^{\mathcal{S}}_F (b_j)) \leq v_i (S) = b_i (S) \). We may assume that VCG prefers to assign larger bundles when tie-breaking between possible assignments. Then, it must be that any allocation to the Sybil bidders is given to one Sybil bidder as a whole bundle. It is then weakly better to send only \( b_j \), as a single bid instead of \( b_i \). Furthermore, it is then weakly better to send \( truth_i \) since truthfulness is dominant for single bid VCG. Since this is true given any nature state, the Sybil attack is weakly dominated by \( truth_i \), which implies \( truth_i \) is loss-averse with respect to it.

This covers all the exact-bidding Sybil attacks. Loss-aversion with respect to overbidding and underbidding attacks are implied by the relevant discussion. Overall this covers all Sybil attacks. \( \square \)

We comment on how to generalize this proof to the mixed loss-averse case. Consider some mixed bid \( b_j \). If there is an overbid in the support, then use the construction of nature state for the overbidding case: Any bid in the support will have a non-positive utility, and the overbid that the nature state is chosen for has negative utility: I.e., overall negative utility in expectation. If there is an underbid in the support, the main issue is this: Truthfulness still performs well, i.e., has a positive utility for any actualized bid in the support, if that bid has non-zero utility, and this positive utility is lower bounded by the grid size \( g \). Moreover, the underbid we construct the nature state for, has 0 utility when actualized. However, the construction needs to be adjusted so that any other actualized bids have utility at most \( g \), which implies the expectation is strictly less than \( g \). The strategy is not loss-averse. The remaining strategies only have exact bids in their support, and so achieve optimal welfare for any actualized bid, and also in expectation. The first part of the proof that \( truth_i \) is loss-averse also requires an adjustment that ensures every actualized Sybil bid has utility of at most \( g \).

5 Discussion and Future Directions

In the example of the discrete first-price auction in Section 2 as well as in our main result in Section 4 the loss-averse solution concept leads to optimal results: truthfulness (or near truthfulness), and optimal revenue or welfare. In Appendix B, we show, by studying different classic settings, that
this is not always the case, and solutions may have various surprising forms.

A robust notion missing from our discussion in Section 4 is min-max regret. We show in Appendix A it does not imply or is implied by our loss aversion notion, and give further characteristics of it. It is also compared with our notion as part of our discussion of voting in Section 5.

In our definition of loss aversion, we consider only pure nature states. We justify this choice in Appendix C by showing that if we consider mixed nature states as well, then the loss aversion and safety level notions become one. In Appendix D we show a particular refinement of our loss aversion notion, and demonstrate why it may be useful.

A few immediate open questions follow our work:

- We find that loss aversion is a stronger notion than safety level. In settings previously studied that proved performance guarantees for safety level strategies, do loss-averse strategies exist? Can they yield better performance guarantees?
- In the case of single-item auctions, our analysis of the discrete first price auction implies that with loss-averse bidders, it is possible to achieve optimal welfare and revenue. Does this extend to combinatorial auctions? If so, does it hold even when the discretization must be polynomially bounded?
- In the presence of partial knowledge or the option to elicitate it (similar to the ideas in [11]), what would the loss-averse action be? This is relevant, for example, when agents arrive sequentially, and so the set of feasible nature states diminishes for later agents.

Acknowledgements

Yotam Gafni and Moshe Tennenholtz were supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant No. 740435).

References

Alkalay-Houlihan, C.; and Vetta, A. 2014. False-Name Bidding and Economic Efficiency in Combinatorial Auctions. Proceedings of the AAAI Conference on Artificial Intelligence, 28(1).

Aumann, R. J. 1985. On the non-transferable utility value: A comment on the Roth-Shafer examples. Econometrica: Journal of the Econometric Society, 667–677.

Babaioff, M.; Dobzinski, S.; and Oren, S. 2018. Combinatorial Auctions with Endowment Effect. In Proceedings of the 2018 ACM Conference on Economics and Computation, EC ’18, 73–90. New York, NY, USA: Association for Computing Machinery. ISBN 9781450358293.

Bell, D. E. 1988. Disappointment in Decision Making under Uncertainty, 358–383. Cambridge University Press.

Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D. 2016. Handbook of Computational Social Choice. USA: Cambridge University Press, 1st edition. ISBN 1107060435.

Camara, M. K. 2022. Computationally Tractable Choice. In Proceedings of the 23rd ACM Conference on Economics and Computation, EC ’22, 28. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391504.

Christodoulou, G.; Kovács, A.; and Schapira, M. 2016. Bayesian Combinatorial Auctions. Journal of the ACM, 63(2).

Chwe, M. S.-Y. 1989. The discrete bid first auction. Economics Letters, 31(4): 303–306.

Cohen, M. D. 1995. Risk-Aversion Concepts in Expected- and Non-Expected-Utility Models, 73–91. Dordrecht: Springer Netherlands. ISBN 978-94-017-2440-1.

Ezra, T.; Feldman, M.; and Friedler, O. 2020. A General Framework for Endowment Effects in Combinatorial Markets. SIGecom Exchanges, 18(2): 38–44.

Feldman, M.; Fu, H.; Gravin, N.; and Lucier, B. 2013. Simultaneous Auctions Are (Almost) Efficient. In Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing, STOC ’13, 201–210. New York, NY, USA: Association for Computing Machinery. ISBN 9781450320290.

Ferejohn, J. A.; and Fiorina, M. P. 1974. The Paradox of Not Voting: A Decision Theoretic Analysis. American Political Science Review, 68(2): 525–536.

Gafni, Y.; Lavi, R.; and Tennenholtz, M. 2020. VCG under Sybil (False-Name) Attacks - A Bayesian Analysis. Proceedings of the AAAI Conference on Artificial Intelligence, 34(02): 1966–1973.

Iwasaki, A.; Conitzer, V.; Omori, Y.; Sakurai, Y.; Todo, T.; Guo, M.; and Yokoo, M. 2010. Worst-Case Efficiency Ratio in False-Name-Proof Combinatorial Auction Mechanisms. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1, AAMAS ’10, 633–640. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9780982657119.

Kleinberg, J.; Mullainathan, S.; and Raghavan, M. 2022. The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimization. In Proceedings of the 23rd ACM Conference on Economics and Computation, EC ’22, 29. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391504.

Kleinberg, J.; and Oren, S. 2018. Time-Inconsistent Planning: A Computational Problem in Behavioral Economics. Communications of the ACM, 61(3): 99–107.

Köblering, V.; and Wakker, P. P. 2005. An index of loss aversion. Journal of Economic Theory, 122(1): 119–131.

Kurokawa, D.; Procaccia, A. D.; and Shah, N. 2018. Leximin allocations in the real world. ACM Transactions on Economics and Computation (TEAC), 6(3-4): 1–24.

Lehmann, B.; Lehmann, D.; and Nisan, N. 2006. Combinatorial auctions with decreasing marginal utilities. Games and Economic Behavior, 55(2): 270–296.

Lu, T.; and Boutilier, C. 2011. Robust Approximation and Incremental Elicitation in Voting Protocols. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume One, IJCAI’11, 287–293. AAAI Press. ISBN 9781577355137.
Merrill, S. 1982. Strategic Voting in Multicandidate Elections under Uncertainty and under Risk. In Holler, M. J., ed., Power, Voting, and Voting Power, 179–187. Heidelberg: Physica-Verlag HD. ISBN 978-3-662-00411-1.

Nisan, N.; Roughgarden, T.; Tardos, E.; and Vazirani, V. V. 2007. Algorithmic Game Theory. Cambridge University Press.

Rabin, M. 2000a. Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion. Department of Economics, Working Paper Series q61d7b4pg, Department of Economics, Institute for Business and Economic Research, UC Berkeley.

Rabin, M. 2000b. Risk Aversion and Expected-utility Theory: A Calibration Theorem. *Econometrica*, 68(5): 1281–1292.

Rabin, M. 2013. Incorporating limited rationality into economics. *Journal of Economic Literature*, 51(2): 528–43.

Rabin, M.; and Thaler, R. H. 2001. Anomalies: risk aversion. *Journal of Economic perspectives*.

Savage, L. J. 1951. The Theory of Statistical Decision. *Journal of Economic perspectives*.

Selten, R. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4(1): 25–55.

Shalev, J. 2000. Loss aversion equilibrium. *International Journal of Game Theory*, 29(2): 269–287.

Tennenholtz, M. 2001. Rational Competitive Analysis. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'01, 1067–1072. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN 1558608125.

Tennenholtz, M. 2002. Competitive safety analysis: Robust decision-making in multi-agent systems. *Journal of Artificial Intelligence Research*, 17: 363–378.

Tversky, A.; and Kahneman, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4): 297–323.

Von Neumann, J.; and Morgenstern, O. 2007. Theory of games and economic behavior. In *Theory of games and economic behavior*. Princeton university press.

Yokoo, M.; Sakurai, Y.; and Matsubara, S. 2004. The effect of false-name bids in combinatorial auctions: New fraud in Internet auctions. *Games and Economic Behavior*, 46(1): 174–188.

### A Min-max Regret

Another robust solution notion is the min-max regret (Savage 1951). The notion has many uses in voting: (Perejohn and Fiorina 1974) showed it can be used to explain why voters choose to participate in elections, and (Merrill 1982) used it to “resolve” the Gibbard-Satterthwaite impossibility theorem (see, e.g., (Brandt et al. 2016)), by showing that plurality voting (for example) is truthful under this notion. (Lu and Boutilier 2011) showed how when only partial preferences are known, voting rules can use this notion to decide a winner, and design good elicitation schemes.

**Definition A.1.** Regret for an action $a_i$ and nature state $a_{-i}$ given a type $\theta_i$ is

$$\text{Reg}(a_i, a_{-i}|\theta_i) = \max_{a_i'} u(a_i', a_{-i}|\theta_i) - u(a_i, a_{-i}|\theta_i).$$

Max regret for an action $a_i$ given a type $\theta_i$ is

$$\text{Reg}(a_i, a_{-i}|\theta_i).$$

A min-max regret action belongs to

$$\arg\min_{a_i} \max_{a_{-i}} u(a_i', a_{-i}|\theta_i) - u(a_i, a_{-i}|\theta_i).$$

In words, the regret of an action $a_i$ under nature state $a_{-i}$ is the maximal lost utility $u(a_i', a_{-i}|\theta_i) - u(a_i, a_{-i}|\theta_i)$ of choosing $a_i$, instead of $a_i'$, over all possible actions $a_i'$ (this regret is non-negative, as there is always the option of choosing $a_i$ itself). Max regret is the maximal such regret over all nature states, and the min-max regret action is the action $a_i$ that has minimal max regret.

We now see that min-max regret yields a different solution to the discrete first-price auction than loss aversion, i.e., the two notions do not imply each other. (Tennenholtz 2001) previously applied min-max regret in auction settings, and in particular discussed the DFPA in their Claim 3.1, which we restate adapted to our notations:

**Claim A.2.** In the discrete first-price auction, the min-max regret strategy is to bid $\epsilon_{\text{net}}(\frac{v_i}{2})$.

**Proof.** For any bid $b_i$, the maximum regret is either $b_i$ itself (in the case when no other bidders show up and it was possible to bid and pay 0), or $v_i - (b_i + \epsilon)$ (in the case when another bidder bids $b_i$ and the item goes to her). We are thus looking for arg min $b_i \max\{b_i, v_i - b_i - \epsilon\}$, among the feasible bids.

For $b_i < \epsilon_{\text{net}}(\frac{v_i}{2})$, the regret is thus at least

$$\text{Reg}(b_i) \geq v_i - b_i - \epsilon \geq v_i - (\epsilon_{\text{net}}(\frac{v_i}{2}) - \epsilon) - \epsilon = v_i - \epsilon_{\text{net}}(\frac{v_i}{2}).$$

For $b_i > \epsilon_{\text{net}}(\frac{v_i}{2})$, the regret is at least

$$\text{Reg}(b_i) \geq b_i \geq \epsilon_{\text{net}}(\frac{v_i}{2}) + \epsilon \geq \max\{\epsilon_{\text{net}}(\frac{v_i}{2}), \frac{v_i}{2}\} \geq \max\{\epsilon_{\text{net}}(\frac{v_i}{2}), v_i - \epsilon_{\text{net}}(\frac{v_i}{2}) - \epsilon\} = \text{Reg}(b_i).$$

We conclude that $\epsilon_{\text{net}}(\frac{v_i}{2})$ is the min-max regret bid strategy.

**Proposition A.3.** Dominant strategy $\implies$ min-max regret

7This is true under worst-case arbitrary tie-breaking. If tie-breaking is uniformly random between bidders of the same bracket, this is still true as the limiting regret when there are $n \to \infty$ bidders in the same bracket.
Definition B.1. All-pay auction: An agent $\varphi$'s regret strategy, while bidding $b$, is the unique loss-averse strategy as well.

Proof. (Proposition B.4) Consider if all agent types are $\frac{1}{2} - \frac{1}{2n}$. Playing loss-averse strategies, all agents declare 0. The facility is thus located at 0 and the welfare loss is $\left(\frac{1}{2} - \frac{1}{2n}\right) n \in \Omega(n)$.

B.3 Voting

Following (Merrill 1982), consider a voter with ordinal preferences over $n$ candidates $c_1 > \ldots > c_n$, that has some underlying normalized cardinal utilities of $1 = f_1 > \ldots > f_n = 0$, i.e., the utility $u$ of the voter from candidate $c_j$ being chosen, given the voter type $(f_1, \ldots, f_n)$, is $f_j$. We consider two positional scoring rules:

Definition B.6. Positional Scoring Rule (PSR) defines a set of permissible $n$-vectors of points over the candidates, and the voter may choose among them. The winner is the candidate with the highest sum of points across voters, where ties are broken towards the worst candidate $c_n$.

The following are positional scoring rules:

Approval - The voter assigns each candidate 0 or 1.

Plurality - The voter chooses a single candidate to assign 1, all others receive 0.

Recall min-max regret as defined in Definition A.1.

Proposition B.7. (Adapted from Merrill 1982)
The min-max regret strategy is truthful for Plurality (choose the top candidate $c_j$). Depending on the underlying cardinal utilities, the min-max regret strategy in Approval is to approve the $k$ top candidates for some $1 \leq k \leq n - 1$.

We include the proof for Plurality since it could be illuminating in comparison to our later result and proof for loss aversion.

Proof. Recall Definition A.1. The max regret of choosing $c_1$ is $f_2 - f_n$, which happens when there is a tie between $c_2$ and $c_n$ and no other candidates, that is decided towards $c_n$.

Then, voting $c_2$ would have tilted the result to $c_2$. The max regret of choosing any other candidate $c_j$ with $2 \leq j \leq n$ is $f_j - f_n$, which happens when there is a tie between $c_1$ and $c_n$, and no other candidate.

We now present the results for loss aversion with Approval and Plurality:

Proposition B.8. (Pure loss-averse)
For Approval, approving all candidates besides $c_n$ is the unique loss-averse strategy.

For Plurality, voting for any $c_j$ with $1 \leq j \leq n - 1$ is a loss-averse strategy.

(Mixed loss-averse)
For Approval, the result is as with pure strategies. For Plurality, there is a unique mixed loss-averse strategy, which takes the form: $(p_1, \ldots, p_n) \propto \left(\frac{1}{f_1}, \ldots, \frac{1}{f_{n-1}}, 0\right)$, where $p_j$ is the probability to vote for candidate $c_j$.

We use a general characterization to derive the results for pure strategies:

*This tie-breaking is w.l.o.g. for our proofs. What matters is that it is possible to pivot the voting result by adding one more vote to a candidate in certain nature states.
Definition B.9. **Pareto domination** - A n-vector \( v \) Pareto dominates another n-vector \( v' \) if \( \forall 1 \leq j \leq n, v_j \geq v_j' \) and \( \exists j \leq n, v_j > v_j' \). A n-vector is in the Pareto frontier of a set \( S \) if it is not Pareto dominated by any vector in \( S \).

Lemma B.10. Consider the set \( S \) as the set of all possible voting vectors under a positional scoring rule. Let \( f(S) = \{ f(v) \}_{v \in S} \) where \( f(v) = \{ (v_1 - v_n, \ldots, v_{n-1} - v_n) \}_{v \in S} \). A voting strategy for a PSR is pure loss-averse if and only if it is in the Pareto frontier of \( f(S) \).

**Proof.** (Loss-averse \( \implies \) Pareto frontier)

Assume towards contradiction voting with \( v \) is not in the Pareto frontier but is loss-averse. Then, there is some \( v' \) so that \( f(v') \) Pareto dominates \( f(v) \) in \( f(S) \). Thus, it must hold that there is some candidate \( j \) so that \( v'_j - v'_n > v_j - v_n \). If by the other votes \( v_n \) has a \( v_n - v_j \) number of votes advantage over \( v_j \) (and a much higher advantage over the other candidates), voting with \( v \) would lead to \( c_n \) winning, while voting with \( v' \) would lead to \( c_j \) winning. So

\[
\min_{v \in \mathcal{D}(v, v')} u(v, v_j) = f_n = 0.
\]

On the other hand, any time \( c_n \) wins under \( v' \) it must win also under \( v \), and so

\[
\min_{v \in \mathcal{D}(v, v')} u(v', v_j) \geq f_{n-1}
\]

\[
> f_n = 0 = \min_{v \in \mathcal{D}(v, v')} u(v, v_j),
\]

which contradicts the loss aversion condition.

(Pareto frontier \( \implies \) loss-averse)

Assume towards contradiction the strategy \( v \) is not loss-averse, then there is some strategy \( v' \) with

\[
\min_{v \in \mathcal{D}(v, v')} u(v', v_j) > \min_{v \in \mathcal{D}(v, v')} u(v, v_j).
\]

But since \( f(v') \) does not Pareto dominate \( f(v) \), there is some candidate \( j \) so that \( v_j - v_n > v'_j - v'_n \), and this shows that

\[
\min_{v \in \mathcal{D}(v, v')} u(v', v_j) = f_n = 0, \text{ in contradiction.}
\]

Claim B.11. **For Approval with mixed strategies, approving all candidates besides \( c_n \) is the unique loss-averse voting.**

**Proof.** Let \( v \) be the stated voting, and let \( v' \) be some other voting. There is some candidate \( c_j \) with \( 1 \leq j \leq n - 1 \) with the probability \( p_j \) of being approved strictly less than 1. Let \( j \) be the maximal such index. Consider the case \( v_j \) of a tie by other voters between \( c_j \) and \( c_n \) (and all other candidates behind), \( u_j(v_j, v_j) = \mathbb{E}_{\tilde{v}' \sim \nu} [\tilde{u}_j(\tilde{v}', v_j)] = p'_j f_j + (1 - p'_j) f_n = p_j f_j < f_j \). In any nature state and random outcome over the mixed voting \( v' \), if \( c_j \) for some \( j \) wins under \( v \), then it also wins under \( v' \). So

\[
\min_{v \in \mathcal{D}(v, v')} u_j(v, v_j) \geq
\]

\[
f_j > \min_{v \in \mathcal{D}(v, v')} u_j(v', v_j), \text{ and so by the loss aversion condition } v' \text{ is not loss-averse (and } v \text{ is loss-averse w.r.t. } v').
\]

\[
\square
\]

Claim B.12. **For Plurality, voting with \( \left( \frac{p_1}{n}, \ldots, \frac{p_n}{n} \right) \) is the unique loss-averse strategy.**

**Proof.** Let \( v \) be the above mixed vote, and let \( v' \) be some other mixed vote. \( v' \) must have some candidate \( c_j \) with \( 1 \leq j \leq n - 1 \) with \( p'_j = \mathbb{P}_{\tilde{v}' \sim \nu} [\tilde{v}' = c_j] < p_j \). Consider the nature states \( v_j \) where \( c_j \) is tied with \( c_n \) and all other candidates are behind, then \( u(v', v_j) = \mathbb{E}_{\tilde{v}' \sim \nu} [u(\tilde{v}', v_j)] = (1 - p'_j) f_n + p'_j f_j = p'_j f_j < p_j f_j = \frac{1}{N_f} \), where \( N_f = \frac{n!}{(n-1)!} \) (the normalizer \( \frac{1}{N_f} \) makes \( \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \) a probability vector). On the other hand, consider some nature state \( v_j \) where it is possible for some candidate \( c_k \) to win over \( c_n \) (but voting for \( c_k \) is required for that). Then, \( u(v_j, v_j) = \mathbb{E}_{\tilde{v}' \sim \nu} [u(\tilde{v}', v_j)] = p_k f_k = \frac{1}{N_f} \).

Any nature state where this is impossible must have that \( c_n \) wins regardless of the player’s vote, but then \( u(v', v_j) = u(v_j, v_j) \). We conclude that \( \min_{v \in \mathcal{D}(v, v')} u(v, v_j) \geq \frac{1}{N_f} > \min_{v \in \mathcal{D}(v, v')} u(v', v_j), \text{ i.e., } v' \text{ is not loss-averse.}

Since this holds for any \( v' \), it also means that \( v \) is loss-averse.

A few aspects of these results are interesting: First, not only truthful strategies are loss-averse in Plurality. In fact, among the pure strategies, all but voting for the worst candidate are equivalent w.r.t. loss aversion. This channels the mindset of a voter willing to vote to anything to avoid the worst candidate. The behavior in Approval voting is similar, though there it gives a unique loss-averse strategy which approves any candidate besides the worst. In approval, it is a “sincere” vote (following the definition in [Merrill 1982]), in the sense that it is monotone in the ordinal preference over candidates (the voter does not approve less desired candidates over more desired candidates). While allowing almost every strategy (being single-mindedly adverse to the worst candidate), loss-aversion is still more useful than the safety level or leximin criteria: Any strategy satisfies both. For the unique mixed solution of Plurality, the idea of mixing votes under uncertainty makes much sense. However, it exhibits a surprising behavior: It gives more weight to candidates that it prefers less, except for the worst one, to which it gives 0 weight. This sounds very counter-intuitive in the context of voting, but is less surprising in the context of robust/safe planning, where it is common to try and balance out the expected utility of different scenarios. We give a simple example to demonstrate this intuition. In this example, there is a unique randomized safety level (and so by Prop. 3.4 also a unique loss-averse) strategy. The strategy gives more weight to a lower utility action than to a higher utility action (if thought about lexicographically).

Example B.13. Consider two agents with actions \( a, b, A, B \) respectively. Let \( u(a, A) = 1, u(a, B) = u(b, A) = \)}
0, \ u(b, B) = 3. Then the unique safety level strategy is
\[ s = \begin{cases} a & \text{w.p. } \frac{3}{4} \\ b & \text{w.p. } \frac{1}{4} \end{cases}. \]

C Discussion of Randomized Environment

In this section we discuss our choice of defining loss aver-

sion w.r.t. deterministic nature-settings (see Defini-
	
tion 2.1). We claim that choosing differently (i.e., allowing ran-

domized nature-settings) would make the notion redundant w.r.t. safety level strategies.

Proposition C.1. If the loss aversion condition applies to all

nature states \( a_{-i} \), both deterministic and randomized, then an action is loss-averse if and only if it is safety level.

Proof. (Loss-averse \( \iff \) safety level)

We know that being loss-averse w.r.t. deterministic nature states implies being safety level (Prop. 3.4.), and adding more restrictions over the notion of loss aversion maintains that.

(Safety level \( \iff \) loss-averse) Let \( a \) be safety level action, \( \theta \) a type. Let \( L \) be the safety level and max \( u_i \), the maximum utility possible under any strategy and any nature state. Assume towards contradiction \( a \) is not loss-averse, i.e., there is an action \( a' \) so that

\[ \min_{a_{-i} \in \bar{D}_\theta(a,a'\mid \theta)} u_i(a, a_{-i} \mid \theta) < \min_{a_{-i} \in \bar{D}_\theta(a,a'\mid \theta)} u_i(a', a_{-i} \mid \theta). \]  

(3)

We can thus fix some nature state \( \bar{a}_{-i} \) so that

\[ u_i(a, \bar{a}_{-i} \mid \theta) < \min_{a'_{-i} \in \bar{D}_\theta(a,a'\mid \theta)} u_i(a', \bar{a}_{-i} \mid \theta) \leq u_i(a_{-i} \mid \theta). \]

Let \( L \) be the safety level and let \( \bar{a}_{-i} \) with \( \forall a'_{-i}, u_i(a', \bar{a}_{-i} \mid \theta) \leq L \) be a nature state that forces the safety level on any action.

Now consider the class of nature states\(^1\)

\[ a'_{-i} = \begin{cases} \bar{a}_{-i} & \text{w.p. } \epsilon \\ a_{-i} & \text{w.p. } 1 - \epsilon \end{cases}. \]

With \( \epsilon \to 0 \), \( u_i(a', a'_{-i} \mid \theta) \leq L + \epsilon \max u_i \to L \). However, the important property of this class is that \( u_i(a_{-i} \mid \theta) \neq u_i(a', a'_{-i} \mid \theta) \) for any member of the class. That is since \( u_i(a_{-i} \mid \theta) = L \) (as \( a_{-i} \) forces safety level strategies to achieve the safety level). So, \( u_i(a, a'_{-i} \mid \theta) = \epsilon u_i(a_{-i} \mid \theta) + (1 - \epsilon)L \neq \epsilon u_i(a', a'_{-i} \mid \theta) + (1 - \epsilon)L = u_i(a', a'_{-i} \mid \theta) \). This means that \( \min_{a_{-i} \in \bar{D}_\theta(a,a'\mid \theta)} u_i(a', a_{-i} \mid \theta) = L \). But, \( \min_{a_{-i} \in \bar{D}_\theta(a,a'\mid \theta)} u_i(a, a_{-i} \mid \theta) \geq L \), as it is a safety level strategy, in contradiction to Eq. (3).

\( \Box \)

\(^1\)One might wonder if such a class really exists among the randomized nature states: We would usually assume that randomized nature states do not include any random outcome, but only these random outcomes that can be a result of individual random actions by agents (i.e., by independent randomization). However in all the settings that we consider in this paper this class can be explicitly created: In voting, by adding an additional voter that votes for the worst candidate w.p. 1 - \( \epsilon \) or does not vote w.p. \( \epsilon \). In auctions, by adding an additional bidder with a very high bid w.p. 1 - \( \epsilon \) or bid 0 w.p. \( \epsilon \), etc.

D Refining the Loss Aversion Notion

For an illustrative example of why it may be needed to further refine our loss aversion notion, consider the following game:

Definition D.1. The aim-big game: An agent may aim big or small (actions are \( B, S \)). Given a nature state \( f_{-i} \in (0, 1] \cup \{1000\} \): If \( f_{-i} \in (0, 1] \),

\[ u_i(S, f_{-i}) = f_{-i}, u_i(B, f_{-i}) = 0. \]

If otherwise \( f_{-i} = 1000 \), then

\[ u_i(B, f_{-i}) = 1000, u_i(S, f_{-i}) = 1. \]

In words, nature chooses either some small reward in \((0, 1]\), or the big reward \(1000\). If the agent aims big, it can not get a small prize (but can get the big prize, if offered). If the agent aims small, it either gets the small prize, or gets a “capped” big prize (only 1 instead of 1000).

Lemma D.2. Both \( B, S \) are loss-averse in the aim-big game.

Proof. \( \min_{f_{-i} \in \bar{D}(B,S)} u_i(B, f_{-i}) = u_i(B, 1) = 0 \) is loss-averse. However, for any \( \epsilon > 0 \), \( u_i(S, \epsilon) = \epsilon \neq 0 = u_i(B, \epsilon) \), and so \( \inf_{f_{-i} \in \bar{D}(B,S)} u_i(S, f_{-i}) = 0 \), and \( B \) is also loss-averse.

This is a weird outcome because intuitively it is clear that aiming small is the loss-averse action: the player always gets more than when aiming big, unless the outcome is very big, in which case it still gets some utility that is bounded away from 0. This is not resolved by eliminating dominated strategies: Both strategies are undominated. However, the following revision can resolve the issue:

Definition D.3. Let

\[ D^*_\theta(a_i, a_i') = \{ a_{-i} \ s.t. \ u_i(a_i, a_{-i} \mid \theta) < u_i(a_i', a_{-i} \mid \theta) \}. \]

We say that an action \( a_i \) of agent \( i \) is **loss averse** (given a type \( \theta \)) if for any other action \( a_i' \),

\[ \min_{a_{-i} \in D^*_\theta(a, a') \mid \theta} u_i(a, a_{-i} \mid \theta) \geq \min_{a_{-i} \in D^*_\theta(a', a') \mid \theta} u_i(a', a_{-i} \mid \theta). \]

(we consider the minimum over an empty set to be \( \infty \))

In the above example, we have

\[ \min_{a_{-i} \in \bar{D}(B,S)} u_i(S, a_{-i} \mid \theta) = 1 > \min_{a_{-i} \in \bar{D}(B,S)} u_i(B, a_{-i} \mid \theta) = 0, \]

and so \( S \) is the unique loss-averse action. This also enables us to complement the picture of Lemma 3.2.

Definition D.4. A strictly dominated action \( a_i' \) has some action \( a_i \) so that for any nature state \( a_{-i} \), \( u_i(a, a_{-i} \mid \theta) \geq u_i(a', a_{-i} \mid \theta) \), and there is some nature state \( a_{-i} \) so that \( u_i(a_i, a_{-i} \mid \theta) > u_i(a_i', a_{-i} \mid \theta) \).

Lemma D.5. A strictly dominated action is not loss-averse.
**E VCG under Sybil attacks: Additional Results**

**Definition E.1.** We say a valuation $v$ is XOS if there are $k$ additive functions $f_1, \ldots, f_k$ over the item set $M$ so that for any $S \subseteq M$, $v(S) = \max_{1 \leq i \leq k} f_i(S)$.

The XOS valuation class strictly contains sub-modular valuations (Lehmann, Lehmann, and Nisan 2006).

**Example E.2.** Consider three bidders, $A$, $B$, and $C$, and four items, $a, b, c$ and $d$. All bidders are XOS (in particular, $A$ is additive) and their XOS additive functions are given in Table 1.

| Bidder A | a | b | c | d |
|----------|---|---|---|---|
|          | 0 | 0 | 3$\epsilon$ | 3$\epsilon$ |
| Bidder B | 0 | 0 | 0 | 0 |
| Bidder C | $\epsilon$ | $\epsilon$ | 9 | 0 |

**Table 1: XOS functions for Example E.2**

Proof. Consider a strictly dominated action $a_i'$, dominated by $a_i$. Fix some nature state $a'_{-i}$, then

$$\min_{a_{-i} \in D^*_{a_i}(a_i',a_i)} u_i(a_i', a_{-i} | \theta_i) \leq u_i(a_i', a'_{-i}),$$

i.e., some finite number. On the other hand the set $D^*_{a_i}(a_i',a_i)$ is empty, so we assign $\infty$ to the minimum over it.

We show that $b_i$ is loss-averse w.r.t. $\text{truth}_i$. First, since it is not overbidding, it is individually rational. Consider the nature state with a single bid $b'$ that has $b'(a) = \theta(a) + b'(c) = 0, b'(a,b) = 2 - \epsilon$. Then $u_i(\text{truth}_i, b') = \epsilon, u_i(b_1, b') = 2\epsilon$, and so $\min_{b_{-i} \in D^*_{a_i}(b_i, \text{truth}_i)} u_i(\text{truth}_i, b_{-i}) = 0$.

The main idea of showing that it is loss-averse to other Sybil attacks is the following: Any other Sybil attack (that is not a single bid) must either underbid $a$ or $b$, or exact-bid them, in which case it must either underbid $c$ or bid it at $c$.

The example shows loss of welfare. E.g., if we consider another additive bidder with $v(a) = 1.5, v(c) = 2\epsilon$, then all items are allocated to her, generating a welfare of $3 + 2\epsilon$. An optimal allocation would allocate item $c$ to agent $i$ and achieve a welfare of $4$. The example can be extended to show arbitrary loss of welfare due to underbidding.

**Example E.3.** (For non-discretized VCG, underbidding can be loss-averse)

Consider three items $a, b, c$, the combinatorial valuation $v_i(a) = 1, v_i(b) = 1, v_i(a,b) = 2, v_i(a,c) = 1 + \epsilon, v_i(a,b,c) = 2 + \epsilon$, and the Sybil attack $b_i = (b_i, b_i, b_i)$ composed of the additive combinatorial bids $b_i(a) = 1, b_i(b) = b_i(c) = 0, b_i(a,b) = 1, b_i(c) = 0, b_i(a) = b_i(b) = 0$.

We show that $b_i$ is loss-averse w.r.t. $\text{truth}_i$. First, since it is not overbidding, it is individually rational. Consider the nature state with a single bid $b'$ that has $b'(a) = \theta(a) + b'(b) = 0, b'(a,b) = 2 - \epsilon$. Then $u_i(\text{truth}_i, b') = \epsilon, u_i(b_1, b') = 2\epsilon$, and so $\min_{b_{-i} \in D^*_{a_i}(b_i, \text{truth}_i)} u_i(\text{truth}_i, b_{-i}) = 0$.

The main idea of showing that it is loss-averse to other Sybil attacks is the following: Any other Sybil attack (that is not a single bid) must either underbid $a$ or $b$, or exact-bid them, in which case it must either underbid $c$ or bid it at $c$.

The example shows loss of welfare. E.g., if we consider another additive bidder with $v(a) = 1.5, v(c) = 2\epsilon$, then all items are allocated to her, generating a welfare of $3 + 2\epsilon$. An optimal allocation would allocate item $c$ to agent $i$ and achieve a welfare of $4$. The example can be extended to show arbitrary loss of welfare due to underbidding.