Noname manuscript No.
(will be inserted by the editor)

False-Name-Proof Voting with Costs over Two Alternatives

Liad Wagman · **Vincent Conitzer**

Abstract In open, anonymous settings such as the Internet, agents can participate in a mechanism multiple times under different identities. A mechanism is false-name-proof if no agent ever benefits from participating more than once. Unfortunately, the design of false-name-proof mechanisms has been hindered by a variety of negative results. In this paper, we show how some of these negative results can be circumvented by making the realistic assumption that obtaining additional identities comes at a (potentially small) cost. We consider arbitrary such costs and apply our results within the context of a voting model with two alternatives.

Keywords Mechanism design, false-name-proofness, voting, revelation principle, anonymity

1 Introduction

In settings with multiple self-interested agents, the agents often need to make a joint decision even though they have different preferences over the possible outcomes. *Mechanism design* provides techniques for reaching a "desirable" outcome in spite of self-interested behavior. Usually, the focus is on *direct-revelation mechanisms*, where agents report their preferences directly to the mechanism, and a decision is made based on these reported preferences. One issue is that agents may have an incentive to misrepresent their preferences. A direct-revelation mechanism is *strategy-proof* if no agent ever has an incentive to misreport. A key result known as the *revelation principle* (Gibbard, 1973; Green and Laffont, 1977; Myerson, 1979, 1981) shows that there is no loss in restricting attention to strategy-proof mechanisms: roughly, for

Early versions of this work were presented at the Twenty-Third Conference on Artificial Intelligence (AAAI 2008), where it was awarded one of two Outstanding Paper Awards; at the International Industrial Organization Conference (2009); and at the University of Chicago Institute on Computational Economics (2009). Parts were also presented at the INFORMS Annual Meeting (2008) and at the World Congress of the Game Theory Society (2008). We thank the participants at these conferences and at the Duke University Microeconomic Theory workshops; the Rice University Microeconomic Theory workshops; and Kyushu University in Japan. We especially thank Makoto Yokoo's research group, Hervé Moulin, Lirong Xia, and anonymous referees for helpful feedback and discussions. Wagman has benefited from support from the Program for Advanced Research in the Social Sciences and from the Yahoo! Faculty Research and Engagement Program. Conitzer is grateful for support through NSF CCF-1101659, IIS-0812113, and IIS-0953756, ARO W911NF-12-1-0550 and W911NF-11-1-0332, and an Alfred P. Sloan Research Fellowship.

Liad Wagman
Illinois Institute of Technology, Stuart School of Business, 565 W Adams St, Suite 452, Chicago, IL 60661, USA.
Tel: (919) 323-6647, Fax: (312) 906-6500, E-mail: lwagman@stuart.iit.edu

Vincent Conitzer
Duke University, Departments of Computer Science and Economics, Levine Science Research Center, Box 90129, Durham, NC 27708, USA. Tel: (919) 660-6503, Fax: (919) 660-6519, E-mail: conitzer@cs.duke.edu

any mechanism (direct revelation or not) which leads to desirable outcomes under dominant strategies, there is a strategy-proof mechanism that leads to equally desirable outcomes.

In a general *social choice* setting, there is a (usually finite) set of alternatives, and agents report ordinal preferences over these alternatives. For this general setting, some negative results are known: for example, the Gibbard-Satterthwaite theorem (Gibbard, 1973; Satterthwaite, 1975) states that if there are at least three alternatives and preferences are unrestricted, then there exists no deterministic mechanism that is nondictatorial (not always selecting the most-preferred alternative of a fixed voter), onto (for every alternative, there are some votes that make that alternative win), and strategy-proof. Gibbard (1977) defines a *decision scheme* to be a function whose domain is the set of preference profiles and whose range is the set of probability distributions over the alternatives. He showed that a strategy-proof decision scheme must be a convex combination of *duple* (assigning positive probabilities to at most two alternatives that are independent of the profile of preferences) and *unilateral* (taking only the preferences of a single voter into account) decision schemes. The negative conclusion of the Gibbard-Satterthwaite theorem is reinforced by Gibbard (1977, 1978): if a mechanism is strategy-proof and Pareto dominated alternatives are never selected, then the scheme must be a random dictatorship.

However, traditional mechanism design assumes that the mechanism can identify every agent. This assumption is not warranted in open, anonymous environments such as the Internet, where it is easy to participate in the mechanism under multiple identifiers (*e.g.*, e-mail addresses). For instance, in the election for *"The Official New 7 Wonders of the World,"*[1] which took place on July 7, 2007, voters could vote by phone, by sending a text message, or on the dedicated election website. Voters could submit multiple votes online using multiple email addresses.

Technological advances over the past two decades have led to the proliferation of consumer review platforms, where users can vote and rate various services and products. These include reviews for hotels (e.g., Hotels.com, Orbitz.com, Priceline.com, Hotwire.com), cruise ships (e.g., Cruisecritic.com, Expedia.com), airlines (e.g., Airlinequality.com), various products (e.g., Amazon.com), experiences and services (e.g., Yelp.com), sellers (e.g., eBay.com), and answers (e.g., Quora.com). On all of these platforms, users may be able to submit (with some effort) multiple reviews. For sellers, boosting the ratings of their own products and tarnishing those of their competitors can provide strong incentives for doing so (Dellarocas, 2006). The recent surge in Question and Answers (Q&A) platforms (including Quora, Amazon's Askville, StackOverFlow, Yedda, Wondir, Yahoo! Answers, among others) has highlighted this issue. To increase direct traffic (as well indirect traffic from improving rankings in search-engine results), website operators would submit answers to questions that relate to their businesses, and subsequently vote in favor of their answers, multiple times. In response, Quora, for instance, has altered its rules of conduct to explicitly forbid its users from voting multiple times in favor of an answer.[2]

1 See `http://www.vote7.com/n7w/world`.

2 See, for instance, `http://techcrunch.com/2011/01/09/quora-4/` and `http://goo.gl/7JWyXE`.

Recent works have shown that reviews on sites such as Yelp.com and Amazon.com substitute for more traditional forms of competition and interact with firms' profits (Chevalier and Mayzlin, 2006; Luca, 2011; Anderson and Simester, 2013). It has been suggested that manipulations of reviews are distorting (Byers *et al.*, 2012) and inefficient (Dellarocas, 2006; Dellarocas and Wood, 2008), both because they entail noisier information and because of the wasteful effort put into the manipulation itself (a phenomenon referred to as a "rat race" among sellers — a situation along the lines of a prisoner's dilemma game). The rating mechanisms used on review platforms, as well as recent attempts to improve them (e.g., Dai *et al.*, 2012), do not fully account for users' ability to participate multiple times. Amazon.com and Yelp.com currently attempt to tackle the problem of multiple and fake participations by creating a reputation system, where reviews by users whose past contributions were deemed valuable by other users receive more weight (e.g., their submissions are listed in the first few reviews). However, multiple participations can occur under such systems as well (for instance, by voting with additional identifiers to increase one's own reputation). Some recent works (Resnick and Sami, 2007, 2008a,b) have proposed to limit the influence of manipulating reviewers in rating mechanisms by tossing away some of the reviews, at the cost of losing some truthful reviews. Our work offers an alternative approach where votes are not thrown away; instead, our method seeks to replicate the majority outcome as closely as possible, while limiting the influence of any one vote to ensure that in equilibrium no fake identifiers are used. Our study of false-name-proof mechanisms also allows one to trade off the costs and benefits of verifying users.[3]

While it is true that individual elections and ratings on platforms such as the above are not as important as large political elections, if one considers all of the different products, services, articles, videos, award nominees, etc., for which users vote and rate on the Internet, their combined worth is significant (e.g., as measured by a percentage of the combined worth of Internet companies that fundamentally rely on such votes).

If the mechanism is such that there is never an incentive for an agent to participate multiple times, the mechanism is said to be *false-name-proof* (Yokoo *et al.*, 2001; Yokoo, 2003; Yokoo *et al.*, 2004; Iwasaki *et al.*, 2010).[4] Again, the restriction to false-name-proof mechanisms is justified by a revelation principle (Yokoo *et al.*, 2004). Unfortunately, the false-name-proofness constraint severely limits the possibilities for mechanism design, as illustrated by a variety of negative results. In combinatorial auctions, no false-name-proof mechanism allocates resources efficiently (Yokoo *et al.*, 2004); later results show that false-name-proofness is impossible even under a weaker maximality constraint (Rastegari *et al.*, 2007) and give bounds on efficiency (Iwasaki *et al.*, 2010). In general voting settings, the situation is even more severe: unless there is unanimity among the voters on some pair of alternatives, any

[3] Conitzer *et al.* (2010) explicitly model optional costly verification of selected users in the context of social networks.

[4] As is the case for strategy-proofness, false-name-proofness is a dominant-strategies criterion, that is, using only one identifier should be optimal regardless of one's preferences and regardless of what the other agents do. A (weaker) Bayesian definition can also be given, but the dominant-strategies version is the one we study in this paper. For an overview of false-name-proofness, see Conitzer and Yokoo (2010).

false-name-proof mechanism (that also satisfies some other minor conditions) must choose the winner uniformly at random (Conitzer, 2008). Moulin (2009) studies the related problem of *routing-proofness* in networks, where each agent seeks to connect a pair of vertices, and an agent can potentially lower her cost by pretending to be multiple agents, whose requested paths will result in a connection between the agent's true vertices. A general approach to removing the incentive for agents to use false names is to perform some limited verification of agents' identities (Conitzer, 2007; Conitzer *et al.*, 2010); of course, this results in reduced anonymity.

In this paper, we show how some of the negative results can be circumvented. In particular, we show that under the assumption that obtaining additional identifiers comes at a (possibly small) nonnegative cost, much more positive results can be obtained. In this setting, a mechanism is false-name-proof if no agent has an incentive to use more than one identifier, *when these costs are taken into account*.

We make the following contributions. We consider *identifier-independent* settings, where the set of possible outcomes is finite and does not depend on which agents are present; voting settings are an example. We show that in such settings, using a costly identifier cannot be a dominant strategy. Based on this, we show a revelation principle for such settings, *i.e.*, for mechanisms in which each type has a dominant strategy of identity creation and reporting, there exists an equivalent mechanism with a dominant strategy of declaring only one identifier with one's true type. We then study voting settings with two alternatives, under both a convex cost structure and an arbitrary cost structure for additional votes. We propose a false-name-proof mechanism and show that it is optimal from certain perspectives.

The paper proceeds as follows. We start with some results on identifier-independent settings in Section 2. We begin to apply these results in a voting model in Section 3. We then study the voting model with non-decreasing marginal costs in Section 4, and generalize our results to arbitrary costs in Section 5. Section 6 concludes.

2 Identifier-independent settings

When extending the theory of mechanism design to highly anonymous settings, one issue about which we need to be careful is the following. In a standard mechanism design setting, the set of agents is assumed known, and because of that, the space of possible outcomes is known. For example, in a single-item auction with two bidders, either bidder 1 wins the item, bidder 2 wins it, or nobody wins it—and the bidders pay/receive some money. However, if we do not know the bidders up front, then, in some sense, we do not know the possible outcomes up front.

In some settings, however, this is not an issue: for example, in a voting setting, we do know the set of available alternatives up front. We say that a setting is *identifier independent* if the set of possible outcomes O does not depend on the agents/identifiers present. In an identifier-independent setting, agents never pay or receive money, and never contribute or receive resources. This rules out auctions, exchanges, and even bartering settings. (The mechanism may, however, decide to build some *public* resource based on the reports.) That is, there is a one-way interaction between an agent and the mechanism: the agent reports preferences and then effectively disappears before any decision is made. Voting settings are perhaps the main example of identifier-independent settings, though agents can also report utilities and other information. Variants of voting settings such as rating mechanisms, surveys, *etc.* are identifier independent.

A highly anonymous, identifier-independent mechanism design setting can be described as follows. Every agent has a type $\theta \in \Theta$ that describes her preferences (and any other private information). There is a commonly known valuation function $v : \Theta \times O \to \mathbb{R}$, where $v(\theta, o)$ gives the valuation that an agent with type θ obtains from outcome o. An agent i then participates in the mechanism under a collection of identifiers $\psi_i^1, \psi_i^2, \ldots, \psi_i^{n_i}$, and takes an action $a_{\psi_i^j} \in A$ for each of these identifiers in the mechanism. (Here, A is defined by the mechanism.)

In a *direct-revelation mechanism*, identifiers report types directly, that is, $A = \Theta$. We assume that the names of the identifiers do not matter. Then, an agent's overall action is given by a multiset of reported types $\{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}$. If θ is the multiset of all reported types, the mechanism is defined by a function $f(\theta)$ which returns the chosen outcome in O. We assume f is defined for every number of reported types, that is, f is *open*.

A direct-revelation mechanism is *strategy-proof* if for any multiset of types θ_{-i} (reported by the other agents), for any $\theta_i, \hat{\theta}_i \in \Theta$, it is the case that $v(\theta_i, f(\theta_i, \theta_{-i})) \geq v(\theta_i, f(\hat{\theta}_i, \theta_{-i}))$. In other words, *among* strategies that use only one identifier, reporting one's true preferences is a dominant strategy. In a strategy-proof mechanism, an agent may still have an incentive to use more than one identifier. However, the agent has no incentive to misreport her preferences under any of these identifiers (this is true *if* the setting is identifier independent—this is not true for, say, combinatorial auctions):

Lemma 1 *In an identifier-independent setting, if f is a strategy-proof mechanism, then any strategy in which i reports false preferences (*i.e.*, different from θ_i) under at least one of her identifiers is dominated by a strategy where i reports her true preferences (*i.e.*, i reports $= \theta_i$) under each of her identifiers.*

Proof Suppose i misreports her preferences for identifier j, that is, $\hat{\theta}_i^j \neq \theta_i$. Let $\hat{\theta}_i^{-j}$ be agent i's other reports, and let $\theta_{-ij} = \theta_{-i} \cup \theta_i^{-j}$ be the set of *all* reports other than $\hat{\theta}_i^j$. By strategy-proofness, $v(\theta_i, f(\theta_i, \theta_{-ij})) \geq v(\theta_i, f(\hat{\theta}_i^j, \theta_{-ij}))$, that is, i is better off changing $\hat{\theta}_i^j$ to θ_i. After repeated applications of these comparisons, we end up with a strategy where every one of i's identifiers reports θ_i.

Lemma 1 thus extends the concept of strategy-proofness to an agent's additional identifiers. Another consideration is that an agent can also change the number of identifiers she uses. In this case, we need to consider the cost of identifiers. In general, the cost of using two identifiers is not necessarily equal to two times the cost of using one identifier. For example, an agent may already have one e-mail address, so using one identifier is free; but she may not have a second one, so that she would spend some effort (incur some cost) setting up another account. Based on this reasoning, the following simple cost model might be used: the first identifier is free, and every additional one has a fixed marginal cost k. While reasonable, this model is still restrictive: the user may not have an existing account, the user may get better at setting up accounts with practice (resulting in decreasing marginal costs), *etc.* (We will return to this specific cost function, which can be written as $c(n_i) = k \cdot (n_i - 1)$ for $n_i \geq 1$, at a later point. We shall refer to it as the *linear* cost function.)

Hence, let us consider an arbitrary nondecreasing cost function $c : \mathbb{N} \to \mathbb{R}$, where $c(n)$ is the total cost of using n accounts ($c(0) = 0$). We assume that agent i's *utility* is separable and is given by $v(\theta_i, o) - c(n_i)$. Then $c(t) - c(t-1)$ for $t > 0$ gives the marginal cost for an agent to participate for the t^{th} time.[5]

Now that we have described how utilities depend on the number of identifiers used, we can define voluntary participation and false-name-proofness.

Definition 1 (Voluntary participation) A mechanism satisfies *voluntary participation* if it is always a best response to use at least one identifier, *i.e.*, if for any θ_{-i}, for any θ_i, we have $v(\theta_i, f(\theta_i, \theta_{-i})) - c(1) \geq v(\theta_i, f(\theta_{-i}))$.

Definition 2 (False-name-proofness) A direct mechanism satisfies *false-name-proofness* (in dominant strategies) if it is never a best response to use more than one identifier, *i.e.*, if for any θ_{-i}, for any θ_i and $\{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}$ (for $n_i \geq 0$), we have $v(\theta_i, f(\theta_i, \theta_{-i})) - c(1) \geq v(\theta_i, f(\{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}, \theta_{-i})) - c(n_i)$.

We note that under this definition, false-name-proofness implies both strategy-proofness and voluntary participation.

Proposition 1 *Consider an identifier-independent setting where for every type θ_i, there is a finite least upper bound $U_{\theta_i} = \sup_{o \in O} v(\theta_i, o) < \infty$ on the utility that type can achieve. Then, for any mechanism f (that is open,* i.e.*, defined for any number of reported types), it is never a dominant strategy for an agent to submit a costly vote (use n_i such that $c(n_i) > 0$).*

Proof For any mechanism f, let O_f be the outcomes in the range of f. We let $U^f_{\theta_i} = \sup_{o \in O_f} v(\theta_i, o) \leq U_{\theta_i} < \infty$ (the highest utility the agent can hope for, given the mechanism f). Suppose in a dominant strategy, type θ_i uses some n_i with $c(n_i) > 0$. We know there exists some θ such that $v(\theta_i, f(\theta)) \geq U^f_{\theta_i} - c(n_i)/2$. However, we then have $v(\theta_i, f(\theta)) - c(0) \geq U^f_{\theta_i} - c(n_i)/2 > v(\theta_i, f(\{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}, \theta)) - c(n_i)$; that is, the agent is better off not participating if the other agents report θ — a contradiction.

Proposition 1 immediately leads to the following corollary.

Corollary 1 *Under the conditions of Proposition 1, if $c(1) > 0$, then no mechanism satisfies voluntary participation.*

Hence, from here on, we assume $c(1) = 0$. The intuitive interpretation of this assumption is that all agents already own an identifier (*e.g.*, an account), so their first participation is costless.[6]

[5] This notation suggests that all agents have the same cost function and that there is no uncertainty about this cost function. However, these assumptions are not necessary: all of our analysis goes through if $c(t)$ is the greatest lower bound on all realizable total costs for obtaining t identifiers.

[6] As is common in the literature, voluntary participation fails in our model if there is a positive cost for the first identifier. However, voting more than once would still be a dominated strategy. That is, given that an agent is going

We now give a revelation principle for false-name-proofness with costs: under certain conditions, if, in an environment where false-name manipulations with costs are possible, a mapping from type vectors to outcomes (a *social choice function*) can be obtained as the result of all agents playing dominant strategies in a mechanism (*i.e.,* the mapping is *implementable in dominant strategies*), then the same mapping can be achieved by a false-name-proof mechanism. We observe that the "traditional" revelation principle only ensures that under f, an agent has a dominant strategy in the model where each agent can report only once. But this does not guarantee that this strategy remains dominant when we extend the game with additional (false-name reporting) strategies.

Theorem 1 *Consider an identifier-independent setting where for every type θ_i, there is a finite least upper bound $U_{\theta_i} = \sup_{o \in O} v(\theta_i, o) < \infty$ on the utility that type can achieve. If $c(1) = 0$ and $c(2) > 0$, then if a social choice function is implementable in dominant strategies by a mechanism g in an environment where false-name manipulations with costs are possible, then it is also implementable by a false-name-proof (direct) mechanism f.*

Proof In the dominant strategies under g, no agent uses more than one identifier, based on Proposition 1 and $c(2) > 0$. We can apply the "traditional" revelation principle to g to obtain a direct-revelation mechanism f that is strategy-proof and satisfies voluntary participation. Suppose, under f, there is some situation where agent i benefits from using more than one identifier. In other words, for some $\theta_i, \theta_{-i}, \{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}$ with $n_i > 1$, we have $v(\theta_i, f(\theta_i, \theta_{-i})) < v(\theta_i, f(\{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}, \theta_{-i})) - c(n_i)$. Let $s(\theta)$ return the dominant strategy for type θ under the original mechanism g (this dominant strategy always involves using 0 or 1 identifiers). We have $g(s(\theta_i), s(\theta_{-i})) = f(\theta_i, \theta_{-i})$ and $g(\{s(\hat{\theta}_i^1), \ldots, s(\hat{\theta}_i^{n_i})\}, s(\theta_{-i})) = f(\{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}, \theta_{-i})$. Let n_i' be the total number of identifiers used by the strategies $\{s(\hat{\theta}_i^1), \ldots, s(\hat{\theta}_i^{n_i})\}$; we have $n_i' \leq n_i$. But then, $v(\theta_i, g(s(\theta_i), s(\theta_{-i}))) = v(\theta_i, f(\theta_i, \theta_{-i})) < v(\theta_i, f(\{\hat{\theta}_i^1, \ldots, \hat{\theta}_i^{n_i}\}, \theta_{-i})) - c(n_i) \leq v(\theta_i, g(\{s(\hat{\theta}_i^1), \ldots, s(\hat{\theta}_i^{n_i})\}, s(\theta_{-i}))) - c(n_i')$, contradicting that $s(\theta_i)$ is dominant for θ_i.

The intuition for Theorem 1 is the following. If there is a beneficial false-name manipulation under f, that same false-name manipulation can be performed under g as well using equally many or fewer identifiers. This is because the dominant strategies under g always use at most one identifier. This contradicts the assumption that s is a dominant strategy under g, proving the result.

3 Voting over two alternatives

In this section, we consider a special case of identifier-independent settings: voting over two alternatives, A and B. We assume that each agent strictly prefers one of the alternatives; which

to participate, the dominant strategy would still be to vote only once. One may also say that if the cost of the first identifier is only ε, then behaving truthfully is ε-dominant. We do imagine settings where casting the first vote is relatively easy (*e.g.*, when the agent comes upon the election, the agent is already logged into her existing account). It is also quite possible for the cost of the first identifier to be *negative*, *i.e.*, getting the first account is actually enjoyable. For example, an agent could get a coupon after voting (ideally, a coupon of which it is useless to have more than one copy). Even if there is some effort cost to voting, this may be exceeded by the worth of the coupon, in which case all of our results go through.

alternative is preferred is private information to the agent. We normalize each agent's utility so that the agent receives utility 1 if its favorite alternative is selected, and 0 otherwise. This is without loss of generality given our assumption of separable utility. Votes can be cast either for A or for B. We observe that given these preferences, the majority rule gives the first-best mechanism, that is, it would maximize social welfare if false-name voting were not a concern. Hence, we aim to get as close as possible to the majority rule, under the constraint of false-name-proofness.

The space of outcomes O is the set of distributions over $\{A,B\}$ (we allow for randomized mechanisms). Because of Theorem 1, assuming $0 = c(1) < c(2)$, we can restrict attention to false-name-proof mechanisms. Hence, each agent votes either for A or for B. Let x_A (x_B) be the number of votes for A (B); we call a vector (x_A, x_B) a *profile*. Let $P_A(x_A, x_B)$ ($P_B(x_A, x_B)$) be the probability that A (B) wins given (x_A, x_B). We require $P_A(x,y) = P_B(y,x)$, that is, the rule is *neutral*. (This also allows us to focus on A without loss of generality.)

The following lemma characterizes false-name-proofness in this setting.

Lemma 2 *Given $c(1) = 0$, a (neutral) rule satisfies false-name proofness if and only if for all $x_A, x_B \geq 0$,*

1. $P_A(x_A+1, x_B) - P_A(x_A, x_B) \geq 0$, i.e., *voting for an alternative cannot diminish its probability of being selected, and*
2. $P_A(x_A, x_B) \leq \min_{t \in \{1,\ldots,x_A-1\}} P_A(x_A - t, x_B) + c(t+1)$, *that is, the expected benefit of using additional identifiers to cast more votes in favor of an alternative does not exceed the cost of doing so.*[7]

Such a rule also satisfies voluntary participation and strategy-proofness. We note that we do not need the analogous conditions for B because we require neutrality.

Proof For the "only if" direction: By voluntary participation ($n_i = 0$ in Definition 2), we have $P_A(x_A+1, x_B) \geq P_A(x_A, x_B)$, or equivalently $P_A(x_A+1, x_B) - P_A(x_A, x_B) \geq 0$. By false-name-proofness, for any $t \in \{1,\ldots,x_A-1\}$, we have $P_A(x_A - t, x_B) \geq P_A(x_A, x_B) - c(t+1)$ (otherwise, if the true profile is $(x_A - t, x_B)$, one of the voters who prefers A would be better off casting a total of $t+1$ votes for A)—but this is equivalent to $P_A(x_A, x_B) \leq \min_{t \in \{1,\ldots,x_A-1\}} P_A(x_A - t, x_B) + c(t+1)$. So the conditions are satisfied.

For the "if" direction: Voluntary participation ($n_i = 0$) follows from $P_A(x_A+1, x_B) - P_A(x_A, x_B) \geq 0$ (equivalently, $P_A(x_A+1, x_B) \geq P_A(x_A, x_B)$). Strategy-proofness ($n_i = 1$) follows from two applications of voluntary participation: $P_A(x_A+1, x_B) \geq P_A(x_A, x_B) \geq P_A(x_A, x_B+1)$. The only manipulation left to consider is an agent who uses multiple identifiers. By Lemma 1, we can assume without loss of generality that this agent votes for A with every identifier. Assume that the agent uses $t+1$ identifiers (that is, t additional ones). If this results in the profile (x_A, x_B), she receives a total utility of $P_A(x_A, x_B) - c(t+1)$. If she had voted only once, she would have had a utility of $P_A(x_A - t, x_B)$. But the latter must be at least as large as the former, because $P_A(x_A, x_B) \leq \min_{t \in \{1,\ldots,x_A-1\}} P_A(x_A - t, x_B) + c(t+1)$. It follows that the rule is false-name-proof.

[7] Since agents' utilities are normalized to 1 for their preferred outcome and 0 for the other outcome, the units in this inequality are consistent.

If $c(n) = 0$ for all $n \geq 0$, then the optimal (in a sense to be made precise later) neutral false-name-proof rule that satisfies voluntary participation is the following "unanimity" rule (Conitzer, 2008):

- If one alternative gets all the votes, select it.
- Otherwise, select an alternative uniformly at random.

The disadvantage of this rule is clear: even if one alternative receives 100 votes and the other alternative receives 1 vote, then a coin is flipped to determine the winner. That is, the rule is not very *responsive* to votes. In some sense, the "most" responsive rule is the majority rule, which chooses the alternative that receives more votes (and flips a coin if there is a tie), thereby maximizing the sum of the utilities.[8] However, the majority rule is not false-name-proof. As we will see shortly, when additional identifiers come at a positive cost, there are false-name-proof rules that are more responsive (more like majority) than the unanimity rule above. Our objective is to maximize responsiveness under the constraint of false-name-proofness, *i.e.*, to get as close as possible to majority.

One may wonder how we should compare two rules if one is more responsive for some profiles, and the other is more responsive for other profiles. When the cost function for additional identifiers is weakly convex, this turns out not to matter, because we will find a rule that is *strongly optimal*, that is, most responsive for all profiles. However, for other cost functions, this turns out to be an issue that we will have to address.

3.1 Rule $FNP2$

We now define a false-name-proof rule that, in a sense that will be made precise below, comes as close to the majority rule as possible. One simple intuition for the rule is the following: taking false-name-proofness as a constraint, the rule maximizes the (probabilistic) impact of each additional vote in favor of the majority winner. The "majority winner" here refers to the winner that the majority rule *would have* produced if every agent voted exactly once. Of course, if we actually used the majority rule, agents would likely use false names.

Definition 3 (FNP2) Rule $FNP2$ sets $P_A(x,0) = 1 - P_A(0,x) = 1$ for all $x > 0$; $P_A(x,x) = .5$ for all $x \geq 0$; for $0 < x_B < x_A$, $P_A(x_A,x_B)$ is recursively defined by:

$$P_A(x_A,x_B) = \min_{t\in\{1,\dots,x_A-1\}} \{P_A(x_A - t, x_B) + c(t+1), 1\}$$

and $P_A(x_A,x_B) = 1 - P_A(x_B,x_A)$ for $0 < x_A < x_B$.

Definition 3 utilizes the fact that under a neutral rule, it is sufficient to characterize $P_A(x_A,x_B)$ for all $x_A \geq x_B$. This is because given any $x_A < x_B$, $P_A(x_A,x_B) = P_B(x_B,x_A) = 1 - P_A(x_B,x_A)$.

[8] One can argue about the precise definition of responsiveness. For example, the rule that chooses A if the total number of votes is odd and B otherwise is more "responsive" in the sense that each additional vote changes the outcome. However, such rules violate neutrality and voluntary participation. For our purposes, a rule is most responsive if, given some constraints (neutrality, false-name-proofness), it comes as close as possible to the outcome of the majority rule.

Equivalently, and perhaps more intuitively, rule $FNP2$ can also be described by the following iterative procedure:

Procedure 1 (FNP2)

1. *Set* $P_A(x_A, 0) = 1 - P_A(0, x_B) = 1$ *for all* $x_A, x_B > 0$, *and* $P_A(x, x) = .5$ *for all* $x \geq 0$.
2. *Initialize* $i := 1$. *Repeat:*
   - *For all* $x_A > i$,
     - (a) *Set* $P_A(x_A, i) = \min_{t \in \{1,\ldots,x_A-1\}} \{P_A(x_A - t, x_B) + c(t+1), 1\}$
     - (b) *Set* $P_A(i, x_A) = 1 - P_A(x_A, i)$
   - $i := i + 1$

Figure 1 provides a sketch of the iterative process described in Procedure 1. We illustrate $FNP2$ with the following example.

Example 1 *Suppose that* $c(1) = 0$, $c(2) = .25$, $c(3) = .3$, *and* $c(i) - c(i-1) \geq .25$ *for all* $i > 3$. *Consider the following table, which gives the probability that A wins under* $FNP2$ *for the profiles* (x_A, x_B), $x_A, x_B \leq 5$*:*

5	0	0	.15	.2	.45	.5
4	0	0	.2	.25	.5	.55
3	0	.2	.45	.5	.75	.8
2	0	.25	.5	.55	.8	.85
1	0	.5	.75	.8	1	1
0	.5	1	1	1	1	1
x_B/x_A	0	1	2	3	4	5

Moving along a fixed row (or fixed column), the differences alternate between .25 *and* .05, *until* 1 *(or* 0*) is reached.*

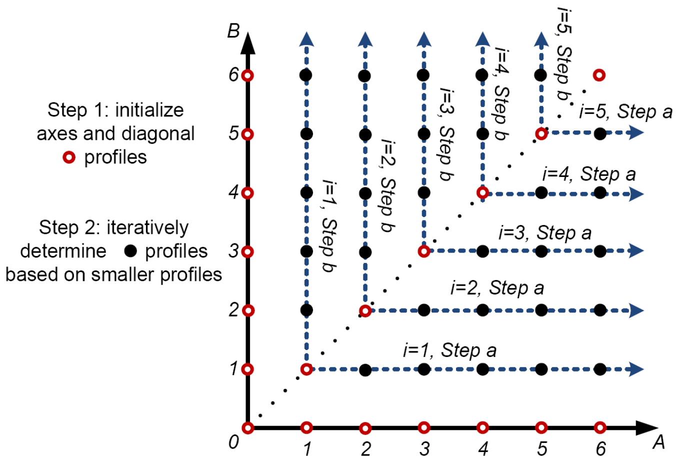

Fig. 1: Sketch of the iterative process in Procedure 1, which describes $FNP2$.

Let $c''(\cdot) \geq 0$ denote a nondecreasing marginal cost of submitting additional votes (obtaining additional identifiers), so that $c(t+1) - c(t) \geq c(t) - c(t-1)$ for all $t > 1$. Similarly, let $c''(\cdot) \leq 0$ denote a nonincreasing marginal cost of additional votes, so that $c(t+1) - c(t) \leq c(t) - c(t-1)$ for all $t > 1$. We note the case $t = 1$ is excluded, so that it is still possible that $c(2) - c(1) > c(1) - c(0)$ even if $c''(\cdot) \leq 0$ (*i.e.*, the first, genuine vote is treated specially, allowing for the situation where agents have "pre-existing accounts"). In fact, this is usually the case because $c(1) = c(0) = 0$. The linear cost model, in which $c(t+1) = k \cdot t$ for $t \geq 0$ and $k > 0$ (so that $k = c(2)$), is a special case of both $c''(\cdot) \geq 0$ and $c''(\cdot) \leq 0$.

Under these conditions, $FNP2$ can be characterized in a simpler way. Cases **(i)** and **(iii)** of the following proposition give the characterizations (Case **(ii)** is an intermediate result which describes the behavior of $FNP2$ in general as long as probabilities of 1 and 0 are not reached at profiles with $x_A, x_B > 0$).

Proposition 2 *FNP2 satisfies:*
(i) *If $c''(\cdot) \geq 0$, the rule FNP2 is primarily defined by the cost of a single additional identifier. In particular, for $x_A \geq x_B$,*

$$P_A(x_A, x_B) = \min\{0.5 + c(2)(x_A - x_B), 1\}$$

(ii) *Let $\pi(\cdot)$ be defined (recursively) as follows: $\pi(1) = 0$ and $\pi(x) = \min_{t \in \{1,\ldots,x-1\}} \{\pi(x-t) + c(t+1)\}$ for $x > 1$. If $P_A(x_A, 1) < 1$, then $P_A(x_A, x_B) = .5 + \pi(x_A) - \pi(x_B)$, that is, the expected benefit to additional votes in favor of an alternative is bounded by the lowest cost of casting them.*
(iii) *If $P_A(x_A, 1) < 1$ and $c''(\cdot) < 0$, then the rule is bounded by the possibility of one agent casting all of the votes in favor of an alternative. In particular,*

$$P_A(x_A, x_B) = .5 + c(x_A) - c(x_B)$$

The proof of Proposition 2 is given in the appendix.

Substituting for $c(\cdot)$ with the linear cost function in Definition 3, $FNP2$ for the linear cost model can be reduced to $P_A(x+t, x) = \min\{1, 1/2 + tk\}$ for $t \geq 0$, $x > 0$, and $k > 0$ (where $P_B(x_A, x_B)$ follows by neutrality, $P_A(x, 0) = 1$ for $x > 0$, and $P_A(0, 0) = .5$).

Lemma 3 *$FNP2$ satisfies voluntary participation and strategy-proofness.*

Lemma 4 *$FNP2$ is false-name-proof.*

The proofs of Lemmas 3 and 4 are relegated to the appendix. To give some intuition for Lemma 3: voluntary participation is equivalent to monotonicity here (it never hurts an alternative to receive another vote), and monotonicity implies strategy-proofness. If $x_A > x_B$, then the probability that A wins when we move from (x_A, x_B) to $(x_A + 1, x_B)$ is nondecreasing because of the recursive formula in the definition of $FNP2$ (combined with the fact that the cost function is nondecreasing). The cases where $x_A = x_B$ and where $x_A < x_B$ are less straightforward. To give some intuition for Lemma 4: by Lemma 1 and by Lemma 3, we can assume without loss of generality that the false-name manipulator only casts votes for her preferred alternative (say, A). If the manipulator's votes result in a profile where $x_A > x_B$, this is not beneficial to the manipulator because of the recursive formula in the definition of $FNP2$: the cost of submitting these additional votes is at least as large their expected benefit. The cases where $x_A = x_B$ and where $x_A < x_B$ are less straightforward.

4 Convex Costs

In this section, we consider the more restrictive case where costs are weakly convex. Such settings make sense when the opportunity cost associated with the effort of obtaining an additional identifier is increasing.[9] We also note that the linear cost model satisfies $c''(\cdot) \geq 0$.

It turns out that for the case of a non-decreasing marginal cost function of additional votes, $c''(\cdot) \geq 0$, where $P_A(x_A,x_B) = \min\{0.5 + c(2)(x_A - x_B), 1\}$, $FNP2$ is *strongly optimal*:

Definition 4 (Strong optimality) A neutral false-name-proof voting rule P is *strongly optimal* if it is closest to the majority rule, that is, if for any other neutral false-name-proof voting rule $\widetilde{P}$, for any profile (x_A,x_B) where $x_A \geq x_B$, we have $P_A(x_A,x_B) \geq \widetilde{P}_A(x_A,x_B)$.

In other words, when $c''(\cdot) \geq 0$, the probability that the alternative with more votes (the majority winner) wins is at least as high under $FNP2$ as under any other false-name-proof rule that satisfies neutrality. We note that a strongly optimal rule is unique by definition.

Theorem 2 *When the marginal cost of additional votes is non-decreasing ($c''(\cdot) \geq 0$), $FNP2$ is the (unique) strongly optimal false-name-proof voting rule with 2 alternatives that satisfies neutrality.*

Proof To prove that $FNP2$ is strongly optimal, it needs to be shown that for any other false-name-proof neutral rule $\widetilde{P}$, for any profile (x_A,x_B) with $x_A \geq x_B$, $P_A(x_A,x_B) \geq \widetilde{P}_A(x_A,x_B)$, where P is $FNP2$. (We recall that the analogous statement when $x_B \geq x_A$ follows by neutrality.) When $c''(\cdot) \geq 0$, $FNP2$ is given by $P_A(x_A,x_B) = \min\{0.5 + c(2)(x_A - x_B), 1\}$. Neutrality requires that for any $x \geq 0$, $\widetilde{P}_A(x,x) = 1/2$. Next, for any $x > 0$, false-name proofness requires that $\widetilde{P}_A(x+1,x) - \widetilde{P}_A(x,x) \leq c(2)$, so that $\widetilde{P}_A(x+1,x) \leq 1/2 + c(2)$. Similarly, $\widetilde{P}_A(x+2,x) - \widetilde{P}_A(x+1,x) \leq c(2)$, so that $\widetilde{P}_A(x+2,x) \leq \widetilde{P}_A(x+1,x) + c(2) \leq 1/2 + 2c(2)$. Continuing in the same manner, for any $t > 0$, $\widetilde{P}_A(x+t,x) \leq 1/2 + tc(2)$ must hold. Also, naturally, $\widetilde{P}_A(x+t,x) \leq 1$. So $\widetilde{P}_A(x+t,x) \leq \min\{1, 1/2 + tc(2)\}$. But $P_A(x+t,x) = \min\{1, 1/2 + tc(2)\}$. Finally, since under $FNP2$, $P_A(x_A,0) = 1$ for $x_A > 0$, clearly $P_A(x_A,0) \geq \widetilde{P}_A(x_A,0)$ holds for all $x_A > 0$. It follows that $FNP2$ is strongly optimal.

5 Arbitrary Costs

We now turn our attention to the more general case where the cost of additional identifiers can be arbitrary. Unfortunately, $FNP2$ is not strongly optimal for general cost functions. The following example illustrates this.

[9] For instance, suppose that obtaining additional identifiers comes at the cost of other activities. If the time value of other activities is concave (*i.e.*, exhibits decreasing marginal value), and if the marginal effort spent to get another identifier is roughly constant, then the opportunity cost per identifier is increasing (as it comes at the cost of ever more valuable alternative activities).

Example 2 *As in Example 1, suppose that* $c(1) = 0$, $c(2) = .25$, $c(3) = .3$, *and* $c(i) - c(i-1) \geq .25$ *for all* $i > 3$. *Consider an alternative rule to* $FNP2$ *over the profiles* (x_A, x_B), $x_A, x_B \leq 5$*:*

x_B/x_A	0	1	2	3	4	5
5	0	0	0	.2	.25	.5
4	0	.2	.2	.45	.5	.75
3	0	.25	.25	.5	.55	.8
2	0	.5	.5	.75	.8	1
1	0	.5	.5	.75	.8	1
0	.5	1	1	1	1	1

As can be seen, the above rule gives a higher probability of A winning at profile $(3,2)$ *than* $FNP2$ *for the same cost specification. On the other hand,* $FNP2$ *gives a higher probability of A winning at* $(2,1)$ *(which is a "smaller" profile).*

The above example suggests that $FNP2$ performs well on small profiles. As we show below, this characterizes $FNP2$ more generally: if we consider it more important to be close to majority on small profiles than on large profiles (for example, because large profiles are less probable), then $FNP2$ is optimal.

We first define a partial order on profiles to make the notion of "smaller" precise.

Definition 5 A profile (x_A, x_B) is said to be *smaller* than another profile (x'_A, x'_B) if $\max\{x_A, x_B\} \leq \max\{x'_A, x'_B\}$, $\min\{x_A, x_B\} \leq \min\{x'_A, x'_B\}$, and at least one of these inequalities is strict. We denote this relationship by $(x_A, x_B) < (x'_A, x'_B)$.

We now define the idea of "being close to majority on small profiles" more precisely.

Definition 6 A false-name-proof neutral rule P is said to be *Most Responsive on Small Profiles (MRSP)* if, given any other false-name-proof neutral rule $\widetilde{P}$ and any profile (x'_A, x'_B) with $x'_A > x'_B$ such that $\widetilde{P}_A(x'_A, x'_B) > P_A(x'_A, x'_B)$, there exists a profile $(x_A, x_B) < (x'_A, x'_B)$ with $x_A > x_B$ where $P_A(x_A, x_B) > \widetilde{P}_A(x_A, x_B)$.

Proposition 3 *If an MRSP rule exists, it is unique.*

Proof Suppose there exist two different MRSP rules, P and $\widetilde{P}$. Let (x'_A, x'_B) be a profile such that $P_A(x'_A, x'_B) \neq \widetilde{P}_A(x'_A, x'_B)$ and such that there is no smaller profile with this property. Without loss of generality, suppose $x'_A > x'_B$ ($x'_A = x'_B$ is not possible because the rules are neutral) and $P_A(x'_A, x'_B) < \widetilde{P}_A(x'_A, x'_B)$. By P's MRSP property, there exists a profile $(x_A, x_B) < (x'_A, x'_B)$ such that $P_A(x_A, x_B) > \widetilde{P}_A(x_A, x_B)$, contradicting the premise that (x'_A, x'_B) is a minimal profile on which the rules differ.

We are now ready to prove that $FNP2$ is the MRSP rule.

Theorem 3 *$FNP2$ is the (unique) MRSP rule.*

Proof Consider any false-name-proof neutral rule different from $FNP2$, denoted by $\widetilde{P}$, and a profile (x_A, x_B) with $x_A > x_B > 0$ at which $\widetilde{P}_A(x_A, x_B) > P_A(x_A, x_B)$. Without loss of generality, suppose this profile is minimal, that is, there is no smaller profile with this property. By the false-name-proofness of $\widetilde{P}$, for any $t \in \{1,\ldots,x_A-1\}$, we must have that $\widetilde{P}_A(x_A,x_B) \leq \widetilde{P}_A(x_A-t,x_B)+c(t+1)$. Because $P_A(x_A,x_B) = \min_{t\in\{1,\ldots,x_A-1\}}\{P_A(x_A-t,x_B)+c(t+1)\}$ and $\widetilde{P}_A(x_A,x_B) > P_A(x_A,x_B)$, it follows that for some $t \in \{1,\ldots,x_A-1\}$, $\widetilde{P}_A(x_A-t,x_B) > P_A(x_A-t,x_B)$. If $x_A-t > x_B$, this contradicts the minimality of the profile (x_A,x_B). If $x_A-t = x_B$, this contradicts the neutrality of the rules. Finally, if $x_A-t < x_B$, we note that $\widetilde{P}_A(x_A-t,x_B) > P_A(x_A-t,x_B)$ is equivalent to $1-\widetilde{P}_A(x_B,x_A-t) > 1-P_A(x_B,x_A-t)$, or $P_A(x_B,x_A-t) > \widetilde{P}_A(x_B,x_A-t)$. The profile (x_B,x_A-t) is smaller than (x_A,x_B), and $x_B > x_A-t$, so we have found a smaller profile on which P is more responsive. Hence, $FNP2$ is the MRSP rule.

6 Conclusions and future work

In open, anonymous settings such as the Internet, an agent can participate in a mechanism more than once without being detected. A mechanism is false-name-proof if no agent ever benefits from participating more than once. In this paper, we considered what happens when there is a cost to participating multiple times. Specifically, we showed that in identifier-independent settings, where the set of possible outcomes is finite and does not depend on which agents are present (such as voting settings), using a costly identifier cannot be a dominant strategy. Based on this, we characterized a revelation principle for such settings. We then studied voting settings with two alternatives and proposed a false-name-proof mechanism that is optimal (where the precise sense of optimality depends on the cost model used).

Future work can take on a number of directions. An immediate direction is to consider weaker (*e.g.*, Bayes-Nash equilibrium) notions of false-name-proofness. Another direction is to generalize our analysis in the domains we studied. This would include studying false-name-proofness with costs in voting settings with more than 2 alternatives.[10] Yet another direction is to extend our analysis to *group* false-name-proof mechanisms with costs, where agents may share the cost of additional identifiers with other agents (*e.g.*, in a social network).[11]

References

Eric Anderson and Duncan Simester. Deceptive reviews: The influential tail, 2013. Working Paper.

John Byers, Michael Mitzenmacher, and Georgios Zervas. The groupon effect on yelp ratings: a root cause analysis. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, New York, NY, USA, 2012.

Judith Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, 2006.

Vincent Conitzer, Nicole Immorlica, Joshua Letchford, Kamesh Mungala, and Liad Wagman. False-name-proof voting in social networks. In *Workshop on Internet and Network Economics (WINE)*, Stanford, CA, 2010.

[10] A preliminary investigation of this setting with linear costs appears in Wagman and Conitzer (2008).

[11] Wagman and Conitzer (2008) introduces group false-name-proof mechanisms in a setting with linear costs. Conitzer *et al.* (2010) study false-name-proof voting (without costs) on social networks.

Vincent Conitzer and Makoto Yokoo. Using Mechanism Design to Prevent False-Name Manipulations. *AI Magazine*, 31(4):65–77, 2010.

Vincent Conitzer. Limited verification of identities to induce false-name-proofness. In *Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 102–111, Brussels, Belgium, 2007.

Vincent Conitzer. Anonymity-proof voting rules. In *Proceedings of the Fourth Workshop on Internet and Network Economics (WINE)*, pages 295–306, Shanghai, China, 2008.

Weijia Dai, Ginger Jin, Jungmin Lee, and Michael Luca. Optimal aggregation of consumer ratings: An application to yelp.com, 2012. Harvard Business School Working Paper, No. 13-042.

Chrysanthos Dellarocas and Charles A. Wood. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 54(3):460–476, 2008.

Chrysanthos Dellarocas. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10):1577–1593, October 2006.

Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41:587–602, 1973.

Allan Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45:665–681, 1977.

Allan Gibbard. Straightforwardness of game forms with lotteries as outcomes. *Econometrica*, 46:595–614, 1978.

Jerry Green and Jean-Jacques Laffont. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, 45:427–438, 1977.

Atsushi Iwasaki, Vincent Conitzer, Yoshifusa Omori, Yuko Sakurai, Taiki Todo, and Makoto Yokoo. Worst-case efficiency ratio in false-name-proof combinatorial auction mechanisms. In *Proceedings of the Ninth International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS)*, pages 633–640, 2010.

Michael Luca. Reviews, reputation, and revenue: The case of yelp.com, 2011. Harvard Business School Working Paper, No. 12-016.

Hervé Moulin. Pricing traffic in a spanning network. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 21–30, Stanford, CA, 2009.

Roger Myerson. Incentive compatibility and the bargaining problem. *Econometrica*, 41(1):61–73, 1979.

Roger Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981.

Baharak Rastegari, Anne Condon, and Kevin Leyton-Brown. Revenue monotonicity in combinatorial auctions. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 122–127, Vancouver, BC, Canada, 2007.

Paul Resnick and Rahul Sami. The influence limiter: provably manipulation-resistant recommender systems. In *RecSys 07: Proceedings of the 2007 ACM conference on Recommender systems*, pages 25–32, 2007.

Paul Resnick and Rahul Sami. The information cost of manipulation-resistance in recommender systems. In *RecSys 08: Proceedings of the 2008 ACM conference on Recommender systems*, pages 147–154, 2008.

Paul Resnick and Rahul Sami. Manipulation-resistant recommender systems through influence limits. *SIGecom Exchanges*, 7(3):1–4, 2008.

Mark Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.

Liad Wagman and Vincent Conitzer. Optimal false-name-proof voting rules with costly voting. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 190–195, Chicago, IL, USA, 2008.

Makoto Yokoo, Yuko Sakurai, and Shigeo Matsubara. Robust combinatorial auction protocol against false-name bids. *Artificial Intelligence*, 130(2):167–181, 2001.

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

Makoto Yokoo. The characterization of strategy/false-name proof combinatorial auction protocols: Price-oriented, rationing-free protocol. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 733–742, Acapulco, Mexico, 2003.

7 Appendix

A Omitted Proofs

LEMMA 3. *FNP*2 *satisfies voluntary participation and strategy-proofness.*

Proof It suffices to show that $P_A(x_A,x_B)$ is weakly increasing in x_A. From Definition 3 and the fact that $c(\cdot)$ is nondecreasing, it immediately follows that for any $0 \le x_B < x_A$, $P_A(x_A+1,x_B) \ge P_A(x_A,x_B)$, because $P_A(x_A,x_B) = \min_{t\in\{1,\dots,x_A-1\}}\{P_A(x_A-t,x_B)+c(t+1),1\} = \min_{t'\in\{2,\dots,x_A\}}\{P_A(x_A+1-t',x_B)+c(t'),1\} = \min_{t'\in\{2,\dots,x_A\}}\{P_A(x_A,x_B)+c(1),P_A(x_A+1-t',x_B)+c(t'),1\} = \min_{t'\in\{1,\dots,x_A\}}\{P_A(x_A+1-t',x_B)+c(t'),1\} \le \min_{t'\in\{1,\dots,x_A\}}\{P_A(x_A+1-t',x_B)+c(t'+1),1\} = P_A(x_A+1,x_B)$.

There are two cases left to prove: **(i)** for $x>1$, $P_A(x+1,x)-P_A(x,x)\ge 0$ (monotonicity at (x,x) profiles; this is not immediately clear from Definition 3 because $P_A(x,x)=.5$ is assigned separately from the recursion) ; and **(ii)** for $x_A<x_B$, $P_A(x_A+1,x_B)-P_A(x_A,x_B)\ge 0$ (monotonicity at profiles where $x_A<x_B$). Actually, for this second case, we will prove the equivalent statement: for $x_B<x_A$, $P_B(x_A,x_B+1)-P_B(x_A,x_B) = P_A(x_A,x_B)-P_A(x_A,x_B+1)\ge 0$.

(i) Consider profile $(x+1,x)$. By Definition 3, $P_A(x+1,x) = \min_{t\in\{1,\dots,x\}}\{P_A(x+1-t,x)+c(t+1),1\}$. In addition, for all $t\in\{1,\dots,x\}$, $P_A(x+1-t,x) = 1-P_A(x,x+1-t)$, and $P_A(x,x+1-t) = \min_{k\in\{1,\dots,x\}}\{P_A(x-k,x+1-t)+c(k+1),1\}$. Thus, $P_A(x,x+1-t)\le$

$P_A(x+1-t,x+1-t)+c(t) = .5+c(t)$. But then $P_A(x+1-t,x) = 1-P_A(x,x+1-t) \geq 1-(P_A(x+1-t,x+1-t)+c(t)) = .5-c(t)$. Therefore, $P_A(x+1,x) \geq \min_{t\in\{1,\ldots,x\}}\{.5-c(t)+c(t+1),1\} \geq .5 = P_A(x,x)$. Hence, for all $x>1$, $P_A(x+1,x)-P_A(x,x) \geq 0$.

(ii) We prove this part by induction. First, $P_A(2,2) = 0.5 \leq P_A(2,1) = \min\{.5+c(2),1\}$. This is the base step. Now, hypothesize that for some $k>1$, $P_A(x_A,x_B) \geq P_A(x_A,x_B+1)$ for all $x_B < x_A < k$. By symmetry, this induction hypothesis implies that $P_A(x_B+1,x_A) \geq P_A(x_B,x_A)$ for all $x_B < x_A < k$ (*i.e.*, voluntary participation holds with respect to alternative B in a square of size $k-1$ and a southwest vertex at $(1,1)$ in a 2-dimensional grid with A on the horizontal axis and B on the vertical axis).

Consider profile (k,x_B) where $x_B<k$ (*i.e.*, extending the square diagonal by one grid point). If $k = x_B+1$, $P_A(k,x_B+1) = .5 \leq P_A(k,k-1) = P_A(k,x_B)$ follows directly from part **(i)**. Suppose then without loss of generality that $x_B+1<k$. From Definition 3, $P_A(k,x_B+1) = \min_{t\in\{1,\ldots,k-1\}}\{P_A(k-t,x_B+1)+c(t+1),1\}$ and $P_A(k,x_B) = \min_{t\in\{1,\ldots,k-1\}}\{P_A(k-t,x_B)+c(t+1),1\}$. By the induction hypothesis, for all $t\in\{1,\ldots,k-1\}$, $P_A(k-t,x_B+1) \leq P_A(k-t,x_B)$. Thus, $P_A(k-t,x_B+1)+c(t+1) \leq P_A(k-t,x_B)+c(t+1)$. But then

$$\min_{t\in\{1,\ldots,k-1\}}\{P_A(k-t,x_B+1)+c(t+1)\} \leq$$

$$\min_{t\in\{1,\ldots,k-1\}}\{P_A(k-t,x_B)+c(t+1)\}$$

It follows that $P_A(k,x_B+1) \leq P_A(k,x_B)$, which completes the induction. Consequently, for $x_B<x_A$, $P_B(x_A,x_B+1) \geq P_B(x_A,x_B)$, which proves **(ii)**.

LEMMA 4. *FNP2 is false-name-proof.*

Proof Part 1 of Lemma 2 follows from Lemma 3. In addition, for every profile where $x_A>x_B$, $P_A(x_A,x_B) \leq \min_{t\in\{1,\ldots,x_A-1\}}\{P_A(x_A-t,x_B)+c(t+1)\}$ follows from Definition 3. It remains to prove that for $1<x_A\leq x_B$, $P_A(x_A,x_B) \leq \min_{t\in\{1,\ldots,x_A-1\}}\{P_A(x_A-t,x_B)+c(t+1)\}$ (we actually prove that for $1<x_B\leq x_A$, $P_B(x_A,x_B) \leq$
$\min_{t\in\{1,\ldots,x_B-1\}}\{P_B(x_A,x_B-t)+c(t+1)\}$, which is equivalent).

We begin by considering a profile (x_A,x_B) where $1<x_B<x_A$ (we treat profiles where $x_B = x_A$ next). Consider a profile (x_A,x_B-k), where $k\in\{1,\ldots,x_B-1\}$. For false-name-proofness, we need $P_B(x_A,x_B) \leq P_B(x_A,x_B-k)+c(k+1)$. Note that

$$P_B(x_A,x_B)-P_B(x_A,x_B-k) \tag{1}$$

$$= 1-P_A(x_A,x_B)-(1-P_A(x_A,x_B-k))$$

$$= P_A(x_A,x_B-k)-P_A(x_A,x_B)$$

By Definition 3, we have $P_A(x_A,x_B) = \min_{t\in\{1,\ldots,x_A-1\}}\{P_A(x_A-t,x_B)+c(t+1),1\}$. The case where $P_A(x_A,x_B) = 1$ holds trivially because $P_A(x_A,x_B-k) \leq 1$. Hence, assume without loss of generality that $P_A(x_A,x_B)<1$ and let $t^\star\in\{1,\ldots,x_A\}$ be such that $P_A(x_A,x_B) = P_A(x_A-t^\star,x_B)+c(t^\star+1)$ (*i.e.*, it is the binding constraint of the minimum). Note that due to $P_A(x_A,x_B-$

$k) = \min_{t\in\{1,\ldots,x_A-1\}}\{P_A(x_A-t,x_B-k)+c(t+1),1\}$, $P_A(x_A,x_B-k) \leq P_A(x_A-t^\star,x_B-k)+c(t^\star+1)$ holds. Then

$$P_A(x_A,x_B-k)-P_A(x_A,x_B) \tag{2}$$

$$\leq P_A(x_A-t^\star,x_B-k)+c(t^\star+1)-(P_A(x_A-t^\star,x_B)+c(t^\star+1))$$

$$= P_A(x_A-t^\star,x_B-k)-P_A(x_A-t^\star,x_B)$$

If $x_A-t^\star = x_B$, we can stop at this point, since $P_A(x_A-t^\star,x_B) = P_A(x_B,x_B) = .5$, and $P_A(x_A-t^\star,x_B-k) = P_A(x_B,x_B-k) \leq P_A(x_B-k,x_B-k)+c(k+1) = .5+c(k+1)$. Combining this with (1) and (2), we obtain $P_B(x_A,x_B)-P_B(x_A,x_B-k) \leq c(k+1)$.

If $x_A-t^\star \neq x_B$, we reiterate the above analysis. In particular, similarly to the above analysis, there exists $t^{\star\star} \in \{1,\ldots,x_A-t^\star-1\}$ such that $P_A(x_A-t^\star,x_B) = P_A(x_A-t^\star-t^{\star\star},x_B)+c(t^\star+t^{\star\star}+1)$. Also similarly, $P_A(x_A-t^\star,x_B-k) \leq P_A(x_A-t^\star-t^{\star\star},x_B-k)+c(t^\star+t^{\star\star}+1)$. Thus,

$$P_A(x_A-t^\star,x_B-k)-P_A(x_A-t^\star,x_B)$$

$$\leq P_A(x_A-t^\star-t^{\star\star},x_B-k)-P_A(x_A-t^\star-t^{\star\star},x_B)$$

Applying this process iteratively, we either reach profile (x_B,x_B), in which case the above conclusion applies, or we obtain

$$P_A(x_A,x_B-k)-P_A(x_A,x_B) \leq P_A(1,x_B-k)-P_A(1,x_B) \tag{3}$$

Since $FNP2$ is defined symmetrically, $P_A(1,x_B-k)-P_A(1,x_B) = (1-P_A(x_B-k,1))-(1-P_A(x_B,1)) = P_A(x_B,1)-P_A(x_B-k,1) \leq c(k+1)$, where the inequality follows from $P_A(x_B,1) = \min_{t\in\{1,\ldots,x_B-1\}}\{P_A(x_B-t,1)+c(t+1),1\}$. Combining all of the above observations, we have $P_B(x_A,x_B)-P_B(x_A,x_B-k) \leq c(k+1)$.

It remains to show that for $x>1$ and $k \in \{1,\ldots x-1\}$, $P_B(x,x)-P_B(x,x-k) = .5-P_B(x,x-k) \leq c(k+1)$. Since $P_A(x,x-k) = \min_{t\in\{1,\ldots,x-1\}}\{P_A(x-t,x-k)+c(t+1),1\}$, $P_A(x,x-k) \leq P_A(x-k,x-k)+c(k+1) = .5+c(k+1)$. Thus, $P_B(x,x)-P_B(x,x-k) = .5-(1-P_A(x,x-k)) = P_A(x,x-k)-.5 \leq c(k+1)$. This completes the proof that $FNP2$ satisfies the conditions of Lemma 2, and is therefore false-name-proof.

PROPOSITION 2. *For all $0 < x_B < x_A$, FNP2 satisfies:*
(i) *If $c''(\cdot) \geq 0$ then*

$$P_A(x_A,x_B) = \min\{0.5+c(2)(x_A-x_B),1\}$$

(The linear cost model in which $c(t+1) = k\cdot t$ for $t,k \geq 0$, whereby $c(2) = k$, is a special case.)
(ii) *Let $\pi(\cdot)$ be defined (recursively) as follows: $\pi(1) = 0$ and $\pi(x) = \min_{t\in\{1,\ldots,x-1\}}$ $\{\pi(x-t)+c(t+1)\}$ for $x>1$. If $P_A(x_A,1) < 1$, then $P_A(x_A,x_B) = .5+\pi(x_A)-\pi(x_B)$*
(iii) *If $P_A(x_A,1) < 1$ and $c''(\cdot) < 0$ then*

$$P_A(x_A,x_B) = .5+c(x_A)-c(x_B)$$

Proof **(i)** Consider $0 < x_B < x_A$ and $k, k' \in \{1, \ldots, x_A - 1\}$ such that $k' < k$. By false-name-proofness, $P_A(x_A - k', x_B) \leq P_A(x_A - k, x_B) + c(k - k' + 1)$. Since $c(1) = 0$ and $c''(\cdot) \geq 0$, $c(k+1) \geq c(k-k'+1) + c(k'+1)$.[12] But then $P_A(x_A - k, x_B) + c(k+1) \geq P_A(x_A - k, x_B) + c(k - k' + 1) + c(k'+1) \geq P_A(x_A - k', x_B) + c(k'+1)$. Thus, the false-name-proofness constraint that $P_A(x_A, x_B) \leq P_A(x_A - k, x_B) + c(k+1)$ is already implied by the constraint $P_A(x_A, x_B) \leq P_A(x_A - k', x_B) + c(k'+1)$, where $k' < k$. Since k and k' were arbitrarily chosen in $\{1, \ldots, x_A - 1\}$, it follows that $P_A(x_A, x_B) = \min\{P_A(x_A - 1, x_B) + c(2), 1\}$. Similarly, $P_A(x_A - 1, x_B) = \min\{P_A(x_A - 2, x_B) + c(2), 1\}, \ldots, P_A(x_B + 1, x_B) = \min\{.5 + c(2), 1\}$. Combining these equalities, we obtain $P_A(x_A, x_B) = \min$
$\{0.5 + c(2)(x_A - x_B), 1\}$.
(ii) It is straightforward to check that $P_A(x_A, 1) = \min_{t \in \{1, \ldots, x_A - 1\}} \{P_A(x_A - t, 1) + c(t+1), 1\} = \min\{P_A(1, 1) + \pi(x_A), 1\} = \min\{.5 + \pi(x_A), 1\}$. If $P_A(x_A, 1) < 1$, then
$P_A(x_A, 1) = .5 + \pi(x_A)$. It also follows from Lemma 3 that for any $k < x_A$, $P_A(k, 1) = .5 + \pi(k)$.

Assume $x_A > 2$. By neutrality, $P_A(1, 2) = 1 - P_A(2, 1) = .5 - \pi(2)$. We also have that $P_A(3, 2) = \min\{1, P_A(1, 2) + c(3), P_A(2, 2) + c(2)\}$. However, $P_A(2, 2) = .5 + \pi(2) - \pi(2) = P_A(1, 2) + \pi(2)$. It follows that $P_A(3, 2) = \min\{P_A(1, 2) + \pi(3), 1\} = \min\{.5 + \pi(3) - \pi(2), 1\}$. Similarly, for any $2 < k \leq x_A$, $P_A(k, 2) = .5 + \pi(k) - \pi(2)$ (where $P_A(k, 2) < 1$ follows from Lemma 3). A similar process can be done for $P_A(k, 3)$ for $3 < k \leq x_A$. Specifically, assuming $x_A > 3$, we have $P_A(1, 3) = 1 - P_A(3, 1) = .5 - \pi(3)$, $P_A(2, 3) = 1 - P_A(3, 2) = .5 + \pi(2) - \pi(3) = P_A(1, 3) + \pi(2)$, and $P_A(3, 3) = .5 + \pi(3) - \pi(3) = P_A(1, 3) + \pi(3)$. It then follows that $P_A(k, 3) = P_A(1, 3) + \pi(k) = .5 + \pi(k) - \pi(3)$.

The proof proceeds by induction (the above being the base step). Hypothesize that for $t \in \{1, \ldots, k_B\}$, $k_B < x_B$, and for any $k_A \leq x_A$, $P_A(k_A, t) = .5 + \pi(k_A) - \pi(k_B)$ (where $P_A(k_A, t) < 1$ follows from Lemma 3 and the assumption that $P_A(x_A, 1) < 1$). By the induction hypothesis, for $x < k_B + 1$, $P_A(x, k_B + 1) = 1 - P_A(k_B + 1, x) = .5 + \pi(x) - \pi(k_B + 1)$. In addition, from the definition of $FNP2$, $P_A(k_B + 1, k_B + 1) = .5 = .5 + \pi(k_B + 1) - \pi(k_B + 1)$. It follows that for $k_B + 1 < k_A \leq x_A$, $P_A(k_A, k_B + 1) = .5 + \pi(k_A) - \pi(k_B + 1)$, which completes the induction. Therefore, if $P_A(x_A, 1) < 1$, then $P_A(x_A, x_B) = .5 + \pi(x_A) - \pi(x_B)$.
(iii) For $k > 0$, $c''(\cdot) < 0$ implies that $c(k+1) - c(k) < c(k) - c(k-1) < \ldots < c(2) - c(1) = c(2)$. Now, since $c(1) = 0$, $c(k) = c(1) + c(k)$. By the above inequalities, $c(k+1) < c(1+1) + c(k)$. Similarly, $c(k+2) < c(1+2) + c(k), \ldots, c(k+t_1) < c(1+t_1) + c(k)$. Applying a similar set of inequalities, we can obtain $c(k + t_1 + \ldots + t_m) < c(1 + t_1) + \ldots + c(1 + t_m) + c(k)$. It follows that $\pi(k) = c(k)$.

Consider any $0 < x_B < x_A$ such that $P_A(x_A, 1) <$. By Part **(ii)**, we have $P_A(x_A, x_B) = .5 + \pi(x_A) - \pi(x_B)$. Combining this with the above, we have $P_A(x_A, x_B) = .5 + c(x_A) - c(x_B)$.

[12] To see this, note that $c''(\cdot) \geq 0$ implies $c(t+1) - c(t) \geq c(t) - c(t-1) \geq \ldots \geq c(2) - c(1)$. Now, since $c(1) = 0$, $c(k'+1) = c(1) + c(k'+1)$. By the above inequalities, $c(k'+1+1) \geq c(1+1) + c(k'+1)$. Similarly, $c(k'+1+2) \geq c(1+2) + c(k'+1), \ldots, c(k'+1+(k-k')) \geq c(1+k-k') + c(k'+1)$. The final inequality gives $c(k+1) \geq c(k-k'+1) + c(k'+1)$.