Crowdsourcing and optimal market design*

Bobak Pakzad-Hurson†

December 30, 2018

Abstract

Solutions to many allocation problems crucially rely on the assumption that agents fully know their preferences over objects to be allocated. I present a general crowdsourcing approach for solving mechanism design problems in which important characteristics of objects are imperfectly observed by agents. The designer first solicits reports of object characteristics by agents and assigns each object a characteristic using a quasi-maximum likelihood method. Second, the designer runs an off-the-shelf "full-information" mechanism using the assessed characteristics. To ensure truth-telling incentives, agents are punished when their reports do not match up with the "wisdom of the crowd." Assuming mild conditions on the relative growth rates of agents and objects, I show this approach yields the same allocation as in the full-information case with probability exponentially converging to one in the number of agents, with aggregate worst-case waste (punishment) converging exponentially to zero. Neither the aggregation nor punishment schemes rely on details of the market. Therefore, my approach is the first to generate near-optimal outcomes with high probability in a variety of settings, including two-sided matching markets. I give necessary and sufficient conditions for recovering desirable properties when signal acquisition is endogenous and costly for agents.

JEL Classification Numbers: D47, D82, C78, C72, D72

Keywords: Market design; information aggregation; interdependent values; matching

*I am grateful to Mohammad Akbarpour, Lint Barrage, Isaias Chaves, Zoe Cullen, Piotr Dworczak, Yinghua He, Matt Jackson, Fuhito Kojima, Shengwu Li, Rich McLean, Stephen Nei, Muriel Niederle, Mike Ostrovsky, Andy Postlewaite, Al Roth, Ran Shorrer, Ilya Segal and seminar participants for helpful comments and suggestions.

†Department of Economics, Brown University, 64 Waterman Street, Providence, Rhode Island 02912 USA. Email: bph@brown.edu.

"For the many, who are not as individuals excellent men, nevertheless can, when they have come together, be better than the few best people, not individually but collectively... For being many, each of them can have some part of virtue and practical wisdom, and when they come together, the multitude is just like a single human being, with many feet, hands, and senses, and so too for their character traits and wisdom."

-Aristotle, *Politics III*

1 Introduction

Optimal allocations are often defined according to the true state of the world. Economists frequently argue that the best workers should be matched to the best firms, that objects should be auctioned off to the buyers with the highest values, and that politicians should only vote to enact laws that are beneficial to society as a whole. However, in arguably most markets, agents do not arrive perfectly informed with all relevant information. Imperfect information can mean that agents' preferences are *interdependent* on the information of others. Most of the market design literature assumes away this issue, and imagines agents that are perfectly informed. This assumption is not without benefit; the literature with interdependent preferences is full of negative results.[1] Because of this, comparatively little is known about designing markets in the presence of interdependent preferences caused by imperfect information.

My goal in this paper is to provide a unified *crowdsourcing* framework which reduces the problem of designing large markets with interdependent preferences to that of well-known full-information market design. This framework aggregates the information of agents and approximates desirable outcomes without relying on the specifics of the market or ideal mechanism. Speaking to the importance of these issues, examples of markets that are specifically designed to aggregate the information of its constituents abound.[2]

This is certainly not the first attempt at aggregating the knowledge of many imperfectly informed

[1] Jehiel and Moldovanu (2001) analyze auctions and show that it is impossible to ensure both truthtelling and efficiency when agents receive multidimensional signals (i.e. observe multiple characteristics imperfectly). Jehiel et al. (2006) show that only constant social choice functions can be implemented. More recently, Che et al. (2015) and Fujinaka and Miyakawa (2015) find negative results in housing assignment settings when agents do not fully observe all relevant characteristics of houses.

In two-sided matching markets *stability*–that the proposed allocation cannot be blocked by pairs of agents through rematching with one another, or blocked by a single agent who rejects her partner for her outside option–is the most common desired property. When agents on one side (workers, for example) each have unknown qualities which are observed via signals by agents on the other side (firms), Chakraborty et al. (2010) show that there is generally no mechanism that yields a stable allocation. Liu et al. (2014) define a weaker notion of stability, which may not even guarantee Pareto efficiency. Bikhchandani (2017) shows that when firms cannot differentiate wages across workers, any maximal matching is stable in the sense of Liu et al. (2014). Therefore a "stable" matching may either be impossible to find or be too permissive.

[2] In massive open online courses (MOOCs) it is not feasible for instructors to grade open-ended assignments of thousands of students. As a solution, multiple students are often assigned to grade the work of each of their peers (Piech et al. (2013)). Individual banks may lack sufficient information to determine the optimal interest rate and the London Interbank Offered Rate (LIBOR) is calculated using the reported interest rates of many banks ((Coulter and Shapiro, 2015)). In an effort to optimally allocate proposal funding, the Sensors and Sensing Systems program at National Science Foundation introduced a scheme in 2013 for proposers to rank the merit of each other's proposals (http://www.nsf.gov/pubs/2013/nsf13096/nsf13096.jsp?WT.mc_id=USNSF_25). The latter two of these markets directly incentivize truthful reports by rewarding agents when their assessments match up with those of others, and punishing the agents otherwise. As I discuss below, this is a key feature of the mechanism proposed in this paper.

individuals. Batchelder and Romney (1988) propose a method of crowdsourcing "without an answer key," Linstone and Turoff (2002) introduce the iterative *Delphi method* (named after the famous oracle in Greek mythology), and Prelec (2004)'s *Bayesian Truth Serum* encourages truthfulness by rewarding reports that are "surprisingly common" given the reports of others. These methods are general in that they seek only to aggregate information, and can be used in a variety of settings. However, they do not consider incentives for truthtelling, or how the resolution of uncertainty can affect market outcomes. Ensuring honesty in these mechanisms requires payments that may be large, and therefore costly to administer.[3]

Others have specificially studied truthtelling incentives. In seminal papers, Cremer and McLean (1985), (1988) propose mechanisms to incentivize truthfulness among agents with correlated information by punishing inconsistent reports. However, the punishments they prescribe can be quite large, possibly violating a limited liability condition or causing sizeable waste. A series of papers, McLean and Postlewaite (2002), (2003), (2015), (2017), and Gerardi et al. (2008), shows that individuals can become *informationally small* as the market grows large, that is, each agent's signal is not likely to drastically change the market assessment. In these settings, a market maker can use small distortions to restore incentives and obtain desirable outcomes. These papers deal with interdependence in Walrasian equilibrium, auctions, the Vickrey–Clarke–Groves mechanism, and expert decision making, respectively.

This paper seeks to combine the generality of the first stand of literature while incentivizing agents to be truthful, taking into account preferences over market outcomes as in the second strand of literature discussed. I create a mechanism that achieves a truthtelling equilibrium that approximates full-information optimal allocations as the market grows large. The mechanism is independent of the particulars of markets, and can be applied to a wide variety of settings, including, but not limited to, the markets studied in McLean and Postlewaite (2002), (2003), (2015), (2017), and Gerardi et al. (2008).

My mechanism crowdsources the identification of important characteristics, and punishes agents when their reports differ from the wisdom of the crowd. The identification method I use is accurate under mild conditions, and gives a closed-form representation of the punishment necessary to ensure truthtelling. Under mild conditions, the total waste generated from the punishment *always* converges to zero in the size of the market. This mechanism can be used even when the number of object characteristics to be identified grows in the size of the market, potentially at a much faster rate than the number of agents. The flexibility of having many objects allows for the analysis of markets in which heterogeneous objects are at play, such as matching markets. These points stand in contrast to mechanisms in the existing literature that consider per capita waste, do not provide closed-form descriptions, and keep constant the number of characteristics to be estimated.[4]

[3]Indeed, Prelec (2004) states that he does "not suggest that people are deceitful or unwilling to provide information without explicit financial payoffs." McKee (2014) attributes the idea that finanial incentives to ensure truthtelling in the Bayesian Truth Serum "would be expensive on a large-scale" to Philip Reny. For example, Rigol and Roth (2016) administer the Bayesian Truth Serum to farmers in a field experiment an India. Without incentivization, subjects lie in ways to favor friends and family. By making payments based on truthfulness scores, they show that subjects can be incentivized to be honest. But this comes at high cost; the average subject receives a payment of roughly 14% of the daily profits of a typical business in the region. Weaver and Prelec (2013) also find that financial incentives are necessary to ensure truthtelling.

[4]Hashimoto (2018) also considers approximation in large markets with multidimensional signals, but with a much

Although the main results are formally stated as limit results, the rates of convergence are exponential meaning that a "large market" need not be prohibitively large. Section 2 demonstrates an application of this mechanism to a hypothetical market similar in size to the job market for junior academic economists. The mechanism is both accurate, in that generating a desirable matching happens with very high probability, and efficient, in that punishments to agents are small.

To lay out the results of the paper, let us begin with a thought experiment involving two markets. In both markets, a designer wishes to allocate objects with varying characteristics drawn from a known distribution to a set of agents. All relevant knowledge, including the realization of all objects' characteristics, is common in the first, full-information market. Suppose there is an associated mechanism φ that yields a desirable allocation b^* according to some properties designated by the designer (for example, b^* might be a Pareto efficient or core allocation, etc.). The second market differs from the first only in that certain characteristics of objects are unknown, and instead, each market participant receives independent and identically distributed signals of these characteristics. When and how is it possible to generate an allocation that approximates b^* when preferences are interdependent? If the agents collectively have enough information it is possible for a market maker to achieve allocation b^* with high probability by running mechanism φ and using the "wisdom of the crowds" assessment of object characteristics as if they were known values, as in the first market.

The first question is: Under what conditions does the market collectively have sufficient information to correctly identify the characteristics of all objects with high probability? Using a quasi-maximum likelihood approach that assigns each object a characteristic most consistent with the received signals, I show that as the size of the market grows, agents' joint assessment of characteristics for all objects converges in probability to the truth at an exponential rate when the number of agents grows faster than the natural logarithm of the number of objects. This result does not follow from the Law of Large Numbers, because although increasing the number of agents (and therefore signals) implies that the probability of identifying the characteristic of a single object increases, making the market larger (i.e. more agents) may also mean an increase in the number of objects, therefore increasing opportunities for mistakes. Proposition 1 gives mild, and tight conditions under which *all* objects are correctly identified as the market grows.

The next question to consider is: Will strategic agents always truthfully reveal their information? If not, do incentives to lie disappear in the limit? By way of examples, I show that agents may have incentives to lie, no matter the size of the market. In fact, I show non-pathological cases in which all agents have weakly dominant strategies to give constant reports of characteristics regardless of their actual signals, meaning that no information is aggregated in equilibrium.[5]

I propose a three-stage croudsourcing mechanism $\bar{\varphi}$. In the first stage $\bar{\varphi}_1$, agents report on their signals of object characteristics. Based on the reports, the mechanism assigns a characteristic to each object using the quasi-maximum liklihood method. Then, using these market estimates, the market designer runs the second stage mechanism $\bar{\varphi}_2$, which is identical to the full-information

different approach that does not allow for the dimension of the signals to grow, such as situations in which each object has its own characteristic. Furthermore, problems such as finding a stable, two-sided matching are not solvable via the use of Hashimoto's budget sets. In an object allocation setting, Akbarpour and Nikzad (2017) approximate desirable allocations using a randomization method.

[5]These examples are similar in spirit to an example in McLean and Postlewaite (2003). They show that even when agents become informationally small there still may be incentives to lie.

optimal mechanism φ. Truth-telling is ensured in equilibrium through stage $\bar{\varphi}_3$ by subjecting agents to a constant, monitary, marginal punishment for each incorrect assessment of the characteristic of an object compared to the estimate by the crowd. I give a closed-form for the punishment that leads to truth-telling in equilibrium so long as agents' utilites are bounded in any feasible allocation. Importantly, this punishment does not depend on the original mechanism φ or other specifics of the market. The proposed punishment scheme does not cause much waste; I show that the worst case sum of punishments, which occurs if every agent incorrectly assesses the characteristics of every object, converges exponentially to zero as the market grows. These penalties can be redistributed back to agents, recycling the waste. As an alternative, I show that in markets in which agents consume disjoint sets of objects and payoffs are bounded away from zero, it is possible to enact the punishment without transfers; agents will be truthful if they are punished with small probabilities of withholding their allocations ex-post. This allows the punishment approach to be viable in markets in which using money is undesirable.

One potential difficulty that can arise using mechanism $\bar{\varphi}$ is that sufficiently punishing agents to remove incentives to game the mechanism can lead to situations in which agents lie solely to avoid punishment. When signals are not strong enough to change an agent's prior belief, she will report the prior belief regardless of the signal she receieves in order to minimize the probability of being penalized. I show that the market designer can intentionally coarsen the set of allowed reports by agents to remove incentives to lie, either to game the mechanism or to avoid punishment.[6] Assuming there exist at least two "sufficiently strong" signals, the identification problem is generically no more difficult after the necessary coarsening of reports.

I also consider settings in which signal strength is endogenously selected by agents, and the cost for picking stronger signals is non-decreasing. The market designer's objective is to identify the characteristics of all objects in large markets, while imposing a total cost (including punishment penalities and information acquisition cost) that goes to zero. This means agents must be incentivized both to pick a sufficiently high signal strength and to make truthful reports. As the market grows, the optimal plan from the designer's point of view involves every agent gathering a very weak signal for the characteristics of each object. When the number of objects remains constant, I find that the necessary and sufficient condition to ensure the desirable properties is identical to a condition necessary to ensure costly information acquisition in certain voting and Condorcet Jury models (Theorem 6 of Martinelli (2006)).[7]

The majority of the paper assumes that utility functions of all agents are common knowledge. I relax this assumption and study what happens when agents' preferences are unknown functions of

[6]One real-world example of coarsening reports is by movie review aggregator Rotten Tomatoes, which gives very fine assessments of quality (to the nearest percent) while relying only on binary signals of quality ("fresh" or "rotten") from each reviewer.

[7]These voting models with endogenous information acquisition rely on voluntary information acquisition, and as a result, not all agents will participate. The present framework makes no claim on the structure of utility functions, and so agents may not have incentives to acquire information without punishment. One interpretation of my findings is that in settings with arbitrary utility functions, all agents will gather small amounts of information (but enough to collectively determine the "best candidate") and participate in elections if there is some small cost of non-participation, perhaps a social stigma. For an example of social norms affecting voting behavior, please see Funk (2005). Funk explains that voter participation dropped in some parts of Switzerland following the implementation of a vote-by-mail scheme. Her explanation is that although mail voting was less costly to voters, it also carried less social status.

object characteristics and must be disclosed to the designer. I show that if the original mechanism φ is strategy-proof under common knowledge of object characteristics, that is, truthfully revealing preferences is a weakly dominant strategy, then mechanism $\bar{\varphi}$ has a truthtelling equilibrium (of both preferences and signals) if each object's characteristic is known by at least one (potentially different) agent. In such a situation where manufacturer Sally knows the characteristic of her own product, while rival manufacturers only recieve noisy signals of the characteristic of Sally's good, the designer can *check* the report by Sally with the *balance* of the joint assessment of all other manufacturers. If Sally is sufficiently punished when her report does not match the assessment of the market, she will always truthfully report her true characteristic in equilibrium. Even when the characteristics of objects are not known to anyone, I show that if φ is an ordinal mechanism, one that does not rely on the cardinal values of agents' preferences, then there exists a truthtelling equilibrium with probability approaching one in the game induced by mechanism $\bar{\varphi}$ as the market grows.

Finally, I allow the market designer to convince naive agents, who form a small but constant proporition of agents in the market, to truthfully reveal their signals. I show that the market designer can detect when groups of agents are jointly lying in order to change the assessment of an object. By punishing all agents in this case, the market designer can remove incentives for coalitional lies, strengthening the solution concept to that of strong Bayes-Nash equilibrium.

The rest of the paper proceeds as follows. Section 2 exposits the general approach of this paper by detailing an application of the mechanism to a two-sided matching model. Section 3 lays out the model and preliminary results. Section 4 describes the proposed mechanism and states the main result. Section 5 presents additional results regarding costly information acquisition, unknown utility functions, and coordinated lying. Section 6 discusses the robustness of the mechanism and concludes. Proofs are relegated to the appendix.

2 Application: Matching with Interdependent Values

There are I firms and $N = f(I)$ workers. Each worker n is equally likely to be of either high or low quality, that is, $q_{(n)} \in \{H, L\}$. Each worker n can be matched to at most a single firm and has strict, transitive preferences over firms and remaining unmatched, which I denote by $\succ_n$. Similarly, firm i can be matched to at most a single worker, and has strict, transitive preference relation over workers and remaining unmatched denoted by $\succ_i$. $\succ_i$ potentially depends on the qualities of workers. For reasons that will become clear, I will represent each firm i's preferences $\succ_i$ via a cardinal utility function $v_i\left(n, q_{(n)}\right)$. I normalize $v_i\left(n, q_{(n)}\right) \geq 0$ for all n and all $q_{(n)}$, and the utility that i receives from remaining unmatched to $\nu_i \geq 0$. Initially, I assume that all of the above is known by a market designer.

The designer wishes to create a one-to-one (ex-post) stable matching between workers and firms.[8] A *matching* is a one-to-one mapping $\mu : I \cup N \to I \cup N$ satisfying $\mu(i) \in N \cup \{i\}$, $\mu(n) \in I \cup \{n\}$

[8]The definition of ex-post stability defined in the paragrah is stronger than the versions of stability in Chakraborty et al. (2010) and Liu et al. (2014). Chakraborty et al. (2010) shows the impossibility of exactly ensuring a stable allocation under their weaker notion of stability.

and $\mu(n) = i$ if and only if $\mu(i) = n$, where the outside option of firm i is selected when $\mu(i) = i$ and similarly for workers. A worker n is *acceptable* to firm i if $n \succ_i i$, and similarly, a firm i is acceptable to worker n if $i \succ_n n$. A matching μ is *(ex-post) stable* if:

- (Individual Rationality) for all $i \in I$, $\mu(i) \succ_i i$ or $\mu(i) = i$ and for all $n \in N$, $\mu(n) \succ_n n$ or $\mu(n) = n$, and
- (No blocking pairs) there is no pair $(i, n) \in I \times N$ such that $n \succ_i \mu(i)$ and $i \succ_n \mu(n)$.

Gale and Shapley (1962) propose the (firm-proposing) deferred acceptance mechanism, which operates as follows:

Step 1: Each firm proposes to its favorite acceptable worker, if such a worker exists. Each worker receiving proposals temporarily holds her favorite, acceptable proposing firm, and rejects all others. If none of the proposing firms are acceptable, then the worker rejects all firms.

Step t: Each firm who was rejected in step $t-1$ applies to her next favorite acceptable worker, if any such workers exist. Each worker receiving proposals in this period considers both these firms and any proposals she may have been temporarily holding from previous steps. From this set of firms, she temporarily holds her favorite, acceptable proposing firm, and rejects all others.

The algorithm terminates when either all firms are being temporarily held by a different worker, or firms who are not being held by any worker have already proposed to all acceptable workers. At this point, the final matching is generated by assigning all workers to the firms they are holding, and leaving all others as unmatched. Gale and Shapley (1962) find that this algorithm always terminates in a finite number of steps, always leads to a stable matching, and moreover, that this stable matching is weakly preferred by each firm to any other stable matching. I will denote this matching μ^F.

Now suppose that the designer knows all primitives other than the qualities of workers. Instead, firms independently observe a noisy signal of the quality of each worker: each firm receives a signal of h independently with probability $\theta > \frac{1}{2}$ and a signal of ℓ independently with probability $1-\theta$ for a high quality worker. Similarly, each firm receives a signal of h independently with probability $1-\theta$ and a signal of ℓ independently with probability θ for a low quality worker.

Can the market designer create matching μ^F in this less informed setting? The first result of the paper deals with situations in which the designer directly observes the signals of each firm. First, I specify a "crowdsourcing" identification strategy that is an application to this setting of the general idenfitication strategy laid out in Section 3. If the proportion of firms that receive signal ℓ for a given worker n is greater than $\frac{1}{2}$, the designer identifies n as being low quality. Otherwise, the designer identifies n as being high quality. If the designer expects to correctly identify the qualities of all workers with high probability, then she can run deferred acceptance using the crowdsourced qualities as if they were the true qualities of the workers. Since the qualities are correctly idenfitied with high probability, this will yield the same matching as μ^Falso with high probability.

A necessary condition to correctly identifying all worker qualities is that the number of firms is large, so that there are many signals of each worker. But a standard Law of Large Numbers argument is insufficient. The reason is that as I grows large, $N = f(I)$ also potentially grows large. Therefore, while the Law of Large Numbers tells us that the probability of correctly identifying a single worker goes to 1, the number of workers for whom it is possible to make a mistake is also growing. Therefore, the rate at which the the number of workers grows relative to the number of firms is crucial. The following finding is a corollary of Proposition 1 in Section 3.

FINDING **1** *: Let $N = f(I) < I^{\alpha}$ for some $\alpha > 0$ and all $I > 0$. Then as $I \to \infty$ the probability of correctly identifying the qualities of all $N = f(I)$ workers exponentially converges to 1.*

Can the market designer still generate matching μ^F even if she does not directly observe the signals, and has to solicit reports from firms? The following example shows that simply asking firms to report their signals creates a game in which each firm has a dominant strategy to ignore its signal and make a pre-specified report regardless of market size.

EXAMPLE **1** *Following a common* limited acceptability *assumption in the matching literature, each worker finds only T firms preferable to her ouside option, regardless of the number of firms in the market (Roth and Peranson (1999), Immorlica and Mahdian (2005), and Kojima and Pathak (2009)). Each firm prefers lower index workers (i.e. worker 1 is preferable to worker 2, all else equal) but finds only high quality workers preferable to its outside option. Due to the alignment of firm preferences, there is a unique ex-post stable matching which assigns the lowest index, high quality worker her favorite firm, the second lowest index, high quality worker her favorite of the remaining firms (or her outside option, if she prefers that), and so on, while leaving all low quality workers unmatched.*

In the reporting game described above, each firm i has incentives to report that all workers who prefer her outside option to i are high quality, as this increases the chances that a desirable worker is freed up for firm i because an "incompatible" worker filled a vacancy at a different firm. The main result of Crawford (1991) implies that any strategy that does not list these incompatible workers as high quality is weakly dominated. But this means that there are at most T truthful reports of each worker's quality, so the probability of correctly ascertaining worker qualities does not converge to 1 regardless of market size. Therefore, the probability of generating μ^F goes to 0 as $I \to \infty$.

The main result of this paper deals with how a market designer who has to ability to impose punishments can ensure matching μ^F with high probability. Consider the following mechanism: The designer solicits reports of signals, using them to identify the qualities of workers as discussed above. In order to ensure truthtelling, the designed punishes firms. In particular, the designer charges each firm a constant penalty each time the firm's report for a particular worker is different from the crowdsourced quality for that worker. For example, if firm i reports quality ℓ for worker n, but more than half of all firms report signal h for worker n, then firm i is punished. Theorem 1 in Section 4 shows that when firms' cardinal utility from workers is bounded above by some constant and their overall utility is quasi-linear in money, then a small punishment is sufficient to ensure truthtelling. Therefore, again, the designer can ensure μ^F with high probability.

Finding **2** *: Suppose $N = f(I) < I^{\alpha}$ for some $\alpha > 0$, there exists D such that $v_i(\cdot,\cdot) < D$ for all $i \in I$, and for all $i \in I$, $u_i = v_i(\cdot,\cdot) - p_i$ where p_i is i's penalty. Then there exists a penalty $\rho(\theta, D, I)$ such that there exists a Bayes Nash equilibrium in which all firms truthfully reveal their signals when they are punished $\rho(\theta, D, I)$ when their report disagrees with the assessment of the market. Moreover, as $I \to \infty$, $N \cdot I \cdot \rho(\theta, D, I) \to 0$.*

In words, this says that the constant marginal penalty required to ensure truthtelling becomes so small in the limit as I grows large that an upper bound on the total punishment assigned (if each firm is punished for each report they make) converges to 0. This punishment can be redistributed, for example, by designating one firm as a non-reporter who instead receives all collected payments. Alternatively, under additional assumptions, I discuss how this punishment can be enforced without money, but by a (vanishing) probability of cancelling a firm's matching and leaving it partnerless.

This result is stated in terms of a "large market," but the mathematical basis for the punishment scheme implies that these markets need not be prohibitively large to achieve desired results. Consider back-of-the-envelope calculations regarding a market with $N = 100$ firms, signal accuracy $\theta = .77$, and the maximum that a firm can value a worker is $D =$\$10 million. Using the methodology of this paper, the chances of the collective information of the market incorrectly assessing the quality of a worker is less than one in one million, and the punishment required to ensure truthtelling is roughly \$1 per incorrect assessment.

3 General Model

3.1 Preliminaries

Let there be $I \in \mathbb{N}$ agents and $N \in \mathbb{N}$ objects, such that the endowment of each agent is e_i. Each object $n < N$ is one of $Q \in \mathbb{N}$ possible characteristics forming the set $Q = \{q_1, ..., q_Q\}$. Let the N^{th} object be money which is of a fixed, known characteristic. It is without loss of generality to assume that each object's characteristic is a scalar.[9] I will write $q_{(n)}$ to represent the characteristic of $n < N$. Let B be the set of feasible allocations, where each $b \in B$ is an $I \times N$ matrix in which $b_{i,n}$ denotes agent i's consumption of object n. The utility function of agent $i \in I$ is common knowledge and represented by $u_i(b, q_{(N)})$, where $b \in B$ and $q_{(N)}$ is the vector of object characteristics. Assume for simplicity that the utility of each agent is quasi-linear in money and that each agent's preferences are unchanged by the monetary transfers of her peers, i.e. for all for all $i, j \in I$, $u_i(b, q_{(N)}) = u_i(b', q_{(N)})$ if $b_{j,n} = b'_{j,n}$ for all $n < N$ and $b_{i,N} = b'_{i,N}$. As the results of this paper deal with asymptotics, I use the superscript "k" to denote the index of the market. Finally, assume that there is some $D < \infty$ that bounds the range of utilities any agent can receive in a feasible allocation, regardless of the index of the market, that is,

$$D \geq \sup_k \left[\sup_{i \in I^k, b \in B^k, q_{(N)}} [u_i(b, q_{(N)})] - \inf_{i \in I^k, b \in B^k, q_{(N)}} [u_i(b, q_{(N)})] \right].$$

[9] To see this, suppose that instead there are two characteristics for each object, each of which can be either high or low. That is, let $Q = \{(L, L), (L, H), (H, H), (H, L)\}$. Let $(L, L) = q_1$, $(L, H) = q_2$ and so on.

Combining these preliminaries, define a market $\bar{\mathcal{M}} = \left\{I, N, q_{(N)}, B, \{e_i\}_{i \in I}, \{u_i\}_{i \in I}, D\right\}$.

3.2 Signals

Now suppose that market $\mathcal{M}$ is identical to market $\bar{\mathcal{M}}$ except that $q_{(N)}$ is unknown and instead observed via noisy signals. Note that I allow $N \cap I \neq \varnothing$, meaning that certain "objects" whose characteristics are unknown can themselves be agents. Each $i \in I$ observes a private signal of characteristics for each object. Denote this signal $\hat{q}_{i,n} \sim F\left(\cdot|q_{(n)}\right)$, where $\hat{q}_{i,n} \in Q$ and $F\left(\cdot|q_{(n)}\right)$ is a non-degenerate distribution for all $q_{(n)}$ with full support over Q. Further assume that $F\left(\cdot|q_{(n)}\right) \neq F\left(\cdot|q_{(n')}\right)$ if $q_{(n)} \neq q_{(n')}$ so that objects of different characteristics generate different signal distributions.

3.3 Information aggregation

Consider a setting in which mechanism φ gives a desireable allocation in a full-information market $\bar{\mathcal{M}}$. In this section, I study two main questions. The first: Under what conditions does "the wisdom of the crowd" have enough information to correctly identify all unknown characteristics in market $\mathcal{M}$ (with high probability)? The second: Is there a mechanism $\bar{\varphi}$ that has an equilibrium whose outcome is close to the allocation derived by φ, i.e. $\varphi\left(\bar{\mathcal{M}}\right) \approx \bar{\varphi}\left(\mathcal{M}\right)$?

For each $n \in N$ define a probability space $(Q, \mathcal{B}, p)$ over possible characteristics. To identify characteristics, a market designer can run a quasi-maximum likelihood estimation by comparing an empirical distribution of signals for a worker to the theoretical distributions expected from objects of different characteristics. Formally, let X_1, X_2, ..., X_I be a sample of i.i.d. observations of a random variable distributed according to distribution F and let $\hat{F}_I(x) = \frac{1}{I}\sum_{\ell=1}^{I} \mathbf{1}_{\{X_\ell \leq x\}}$ be the empirical distribution of the sample

Consider the joint assessment of agents of a single object $n \in N$. For any distinct characteristics q_o and q_p define

$$\epsilon^{q_o, q_p} = \frac{\max_{\hat{q}} \left|F\left(\hat{q}|q_o\right) - F\left(\hat{q}|q_p\right)\right|}{2}$$

and let

$$\epsilon = \min_{q_o \neq q_p} \left[\epsilon^{q_o, q_p}\right] \tag{1}$$

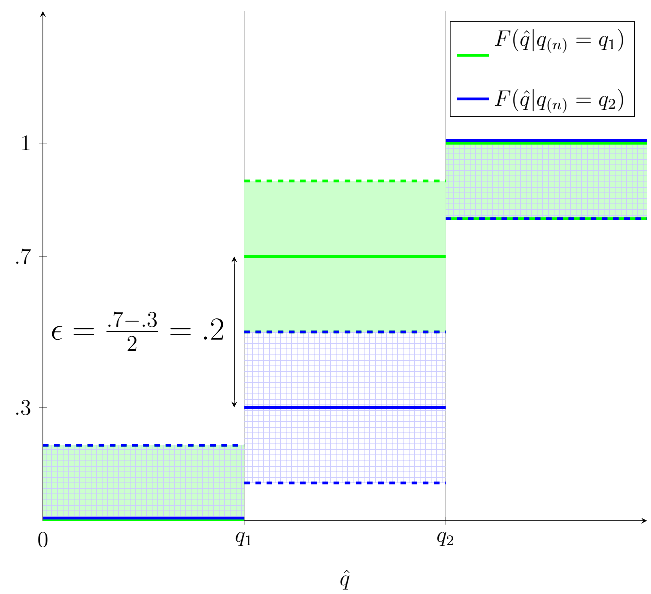
meaning that there is at least 2ϵ distance between the signal distributions of any two objects with different characteristics at some point. Say that the characteristic of $n \in N$ is $\epsilon-$ *identified* as q_m if

$$\left|\hat{F}_I\left(\hat{q}\right) - F\left(\hat{q}|q_m\right)\right| \leq \epsilon \text{ for all } \hat{q}.$$

In other words, we say that an object is identified as having characteristic q_m if the distance between the empirical distribution of signals for said object and the conditional distribution of a q_m object is no larger than ϵ according to the supremum norm,. By construction, for ϵ satisfying Equation (1), an object generically cannot be identified as having two different characteristics. Say that the characteristic of object n is $\epsilon-$*correctly identified* if $\left|\hat{F}_I(\hat{q}) - F\left(\hat{q}|q_{(n)}\right)\right| \leq \epsilon$ for all $\hat{q}$.

To represent this graphically, consider the following illustration with two possible characteritics for each objects. ϵ is defined as above such that an empirical distribution cannot fall within ϵ distance pointwise of both conditional signal distributions.

Figure 1: Identifying Object Characteristics



Notes: Let $Q = \{q_1, q_2\}$. ϵ is chosen so that the maximum distance between $F(\hat{q}|q_1)$ and $F(\hat{q}|q_2)$ is exactly ϵ. Objects whose empirical signal distributions fall in the ϵ band around $F(\hat{q}|q_1)$ (the green, solid region) are identified as q_1. Objects whose empirical signal distributions fall in the ϵ band around $F(\hat{q}|q_2)$ (the blue, checkered region) are identified as q_2.

What are the conditions under which a market designer with access to the signals of all agents can $\epsilon-$correctly identify the characteristics of *every* object with high probability?Clearly, the number of signals on each object must be large, but a standard Law of Large Numbers argument is insufficient in this setting. The reason is that as the market grows large, the probability of correctly identifying the characteristic of a single object goes to 1 (LLN) but the number of objects being assessed may also increase, so there are also more opportunities to make mistakes. Therefore, the rate at which the number of objects grows compared to the number of agents is crucial. To formalize the comparison between the growth rates of different sequences, I provide the following definition.

Definition **1** *:*

- $f^k \in \Omega\left(g^k\right)$ *if there exists* $\delta > 0$ *and there exists* K *such that for all* $k > K$, $f^k \geq \delta \cdot g^k$,
- $f^k \in o\left(g^k\right)$ *if for all* $\delta > 0$ *there exists* K *such that for all* $k > K$, $f^k \leq \delta \cdot g^k$,

- $f^k \sim g^k$ *if for all* $\delta > 0$ *there exists* K *such that for all* $k > K$, $\left|\frac{f^k}{g^k} - 1\right| < \delta$.

Intuitively, $f^k \in \Omega\left(g^k\right)$ when f^k is not asymptotically dominated by g^k, $f^k \in o\left(g^k\right)$ when g^k asymptotically dominates f^k, and $f^k \sim g^k$ when f^k and g^k grow at the same rate asymptotically. As this paper is concerned with asymptotics involving large numbers of agents and objects, when I use this notation, I also implicitly assume that $f^k \to \infty$ and $g^k \to \infty$ as $k \to \infty$.

The following result states that a market designer can correctly identify the characteristics of all objects with high probability if the number of agents of each type grows faster than the natural logarithm of the number of objects, and that this bound is tight when there are more than three possible characteristics.

PROPOSITION **1** : *In a sequence of markets* $\mathcal{M}^1, \mathcal{M}^2, \ldots$

1. *the probability of* $\epsilon-$*correctly identifying every object for* ϵ *satisfying Equation (1) converges to one if* $I^k \in \Omega\left(\log\left(N^k\right)\right)$,
2. *when* $N^k \in o\left(\left(I^k\right)^\alpha\right)$ *for some* $\alpha > 0$ *the rate of convergence is exponential, and*
3. *the probability of* $\epsilon-$*incorrectly identifying a non-vanishing proportion of objects converges to one if* $I^k \notin \Omega\left(\log\left(N^k\right)\right)$ *and* $|Q| > 3$.

One observation is that the market is made large by increasing both the number of agents and objects, not by "cloning" the market (as seen, for example, in McLean and Postlewaite (2002, 2003)). In many contexts, simply duplicating agents to make the market large is undesireable, as "large" markets have both many agents and many objects, and duplicating the preferences of agents in a small market will not lead to rich preferences over all objects. On such example is a two-sided matching market.

3.4 Incentives to manipulate mechanisms

The preceeding section established that identification of the characteristics of all objects is relatively easy under truthtelling as the market grows in size, but do agents always have incentives to tell the truth? If not, do any incentives to lie disappear in the limit? By way of examples, I show that agents may have incentives to lie no matter the size of the market; agents may have simple, dominant strategies to lie, leading to very different outcomes than under truthtelling. These examples are of a two-sided matching market (in which the desired solution concept is stability, see Example 1 in Section 2), a market for selling goods (in which the desired solution concept is a competitve equilibrium, see Example 3 in the appendix), and a voting market (see Example 4 in the appendix). Example 1 demonstrates a situation in which each agent has a dominant strategy to report that nearly objects are of high quality while in Example 3, each agent has a dominant strategy to report that nearly all objects are of low quality. In Example 4, agents are polarized and vote according to party lines. In these examples, it is not feasible to simply ignore the assessments of those who have desires to manipulate because the remaining set of individuals (in the first example, this set is empty) do not have sufficient information to assess the characteristics of all objects.

3.5 How to elicit truthful signals by creating small amounts of waste

The main goal of this paper is to take a full information market $\bar{\mathcal{M}}$, an ideal mechanism φ and a resulting allocation b^* and show how to recover allocation very close to b^* in market $\mathcal{M}$ where characteristics are imperfectly observed instead of common knowledge. Examples 1, 3, and 4 demonstrate that even in situations in which the market has sufficient information to correctly identify the characteristics of all objects, a revelation game in which agents are asked to reveal their signals does not necessarily have a truthtelling equilibrium. Agents may have very simple dominant strategies that result in drastically different outcomes in equilibrium than under truthtelling. To get around this problem, I study a mechanism that has a Bayes-Nash equilibrium in which all agents truthfully reveal signals which approximately achieves the desired allocation b^*. The mechanism uses market reports to assign each object a characteristic, as in Section 3.3. To ensure truthtelling incentives, an agent is charged a constant marginal punishment each time her report for an object's characteristic differs from that of the market assessment.

3.5.1 Limits of punishment

I defer the question of how to implement such punishment, and first study the following: If the punishment scheme is sufficiently stringent so that every agent's equilibrium payoff is maximized by minimizing punishment, when are the reports of the market able to (with high probability) correctly identify the characteristics of all objects? Clearly, any sufficient condition will be no weaker than that in Proposition 1. Nevertheless, the following example demonstrates that the condition can be strictly stronger.

Example **2** : *There are N^k objects and I^k agents, where $I^k = N^k$. Suppose there are two possible characteristics q_1 and q_2. Each object has an ex-ante $\frac{9}{10}$ chance of being q_1. Let the probability that each agent's signal of an object's characteristic is correct be $\frac{3}{4}$. Notice that as $I^k \to \infty$ the probability that each object's characteristic is correctly identified under truth-telling goes to 1 (Proposition 1).*

As an extreme case, suppose that each agent i's utility $u_i = -t_i$ where t_i is the penalty i must pay. In other words, agents do not care about the allocation they receive, they only care about avoiding punishment as much as possible. Notice that $p(q_{(n)} = q_2|\hat{q} = q_2) = \frac{p(\hat{q}=q_2|q_{(n)}=q_2)\cdot p(q_{(n)}=q_2)}{p(\hat{q}=q_2)} = \frac{1}{4} < p(q_{(n)} = q_2|\hat{q} = q_1)$ which means that from i's point of view, regardless of the signal received, any object is more likely to have characteristic q_1. Assuming all other agents are truthfully reporting signals, all objects will be correctly identified with high probability. Then any agent i has an incentive to report q_1 for all objects regardless of her own signal.[10]

The problem with truthful reporting in Example 2 is that agents have incentives to misreport signals solely to avoid punishment. Clearly this can be a problem when agents always wish to report one characteristic for all objects (when others are telling the truth), however, this section

[10] In order for an agent to report q_2 for an object n when it sees signal $\hat{q} = q_2$ it must be the case that $p(n \text{ is identified as } q_2|\hat{q} = q_2) \geq \frac{1}{2}$. But $p(n \text{ is identified as } q_2|\hat{q} = q_2) = p(n \text{ is identified as } q_2|q_{(n)} = q_2, \hat{q} = q_2) \cdot p(q_{(n)} = q_2) + p(n \text{ is identified as } q_2|q_{(n)} = q_1, \hat{q} = q_2) \cdot p(q_{(n)} = q_1)$ $\leq 1 \cdot \frac{1}{10} + p(n \text{ is identified as } q_2|q_{(n)} = q_1, \hat{q} = q_2) \cdot \frac{9}{10}$. Therefore, in order for an agent to be willing to truthfully report $\hat{q} = q_2$ it must be the case that at least $\frac{4}{9}$ of q_1 objects are misidentified.

will show that when there are multiple *hubs,* signals which are reported in equilbrium, and under slight genericity conditions, this does not impose a restriction on the ability of a designer to elicit responses which correctly assay the characteristics of all objects with high probability.

Let

$$r(q) = \underset{q'}{\operatorname{argmax}}\, p(q'|\hat{q} = q)$$

represent an agent's best guess as to the characteristic of an object upon receiving signal q. Define a directed multigraph $G = (Q, E)$ where $E = \{(q, r(q)\}|q \in Q\}$ is the set of directed edges in which each signal q "points to" $r(q)$. Note that it is possible to have $q = r(q)$ implying a loop. Call a node q a *hub* if it has indegree of at least 1. Let H be the set of all hubs. Call those nodes pointing to a hub *siphons* and for any hub q let $s(q) = \{q'|(q', q) \in E \text{ and } q \in H\}$ be its siphons. A node can be both a hub and a siphon, and due to the possibility of loops, it can be the case that $q \in s(q)$.

Punishment has competing effects. On one hand, agents may want to misreport signals in order to game the mechanism. On the other, they may want to lie about signals in order to avoid punishment (i.e. report a hub instead of a siphon). For the rest of the paper, I will consider *direct mechanisms* that never give agents any incentive to lie to avoid punishment. A direct mechanism is Bayesian-Nash-Incentive-Compatible (BNIC) if there exists a Bayesian Nash equilibrium in which agents truthfully report their signals. To remove incentives to lie, the market designer interprets a report that is a siphon to mean that the agent has in fact reported the hub containing said siphon. For example, in Figure 2, the only hubs are q_2 and q_4 and since $q_{1,}q_2, q_4$ and q_3 are the siphons of $q_2$and q_4 respectively, the designer treats any report of q_1, q_2 or q_4 as q_2 and treats a report of q_3 as q_4.

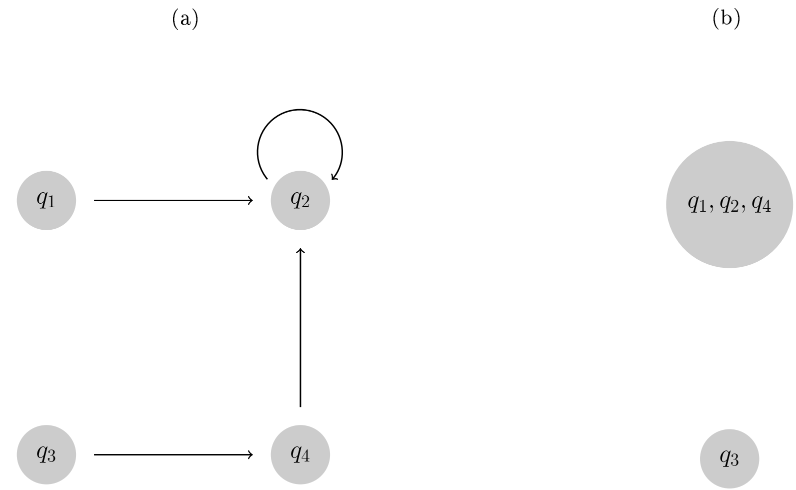
When the number of hubs is equal to the number of possible characteristics, there are no additional complications to identification, as $s^{-1}(q)$ is well defined. But what happens when there are non-trivial hubs and siphons, i.e. $r^{-1}(q)$ is not uniquely defined? For a hub q, it is now not clear if a report of q indicates that the agent received signal q or if it recieved a signal of one of q's siphons.

Because of this, partition Q into sets containing a hub and its siphons for use in characteristic identification, that is $Q = \{\bar{H}^h(q)\}_{h=1}^{|H|}$ where $\bar{H}^h(q) = \{q_h \cup s(q_h)|q_h \in H\}$. Let $\bar{\mathcal{F}}$ be a partition of Q such that for distinct q, q', $q' \in \bar{\mathcal{F}}(q)$ if and only if $p\left(\hat{q} \in \bar{H}^h(q)|q'\right) = p\left(\hat{q} \in \bar{H}^h(q)|q\right)$ for all $\bar{H}^h(q)$. $\bar{\mathcal{F}}$ is a partiton of Q such that members of the same partition have the same empirical distribution over $\bar{H}^h(q)$, which takes into account that the designer does not destinguish between the siphons of each hub. Essentially, the designer is attempting to distinguish Q different characteristics with possibly fewer than Q distinct signal reports.

Another difference to take into account is that the ϵ bound used in identifying object characteristics will generally need to be adjusted to take into account that only a subset of characteristics will be reported by agents in equilibrium. Consider any two characteristics q_o and q_p such that $\bar{\mathcal{F}}(q_o) \neq \bar{\mathcal{F}}(q_p)$. Define

$$\epsilon^{q_o, q_p} = \frac{\max_{\hat{q}} \left|F\left(\hat{q} \in \bar{H}^h|q_o\right) - F\left(\hat{q} \in \bar{H}^h|q_p\right)\right|}{2}$$

Figure 2: Hubs and Siphons



Notes: Panel (a) denotes hubs and siphons for a particular signal structure. The only hubs are q_2 and q_4, with siphons q_1, q_2, q_4 and q_3, respectively. Panel (b) denotes the coarsening of the signal structure that the mechanism undertakes to remove incentives to misreport. For identification purposes, the mechanism treats any report of q_1, q_2 or q_4 as q_2 and treats a report of q_3 as q_4.

and let

$$\epsilon = \min_{q_o, q_p} \epsilon^{q_o, q_p}. \tag{2}$$

ϵ again is formulated such that there is at least 2ϵ distance between the expected reported signal distributions for objects of different characteristics. For ϵ defined in Equation 2, I say that object n is $\epsilon-$ *identified* as having characteristic q_p if $\left|\hat{F}_I\left(\hat{q} \in \bar{H}^h\right) - F\left(\hat{q} \in \bar{H}^h | q_p\right)\right| \leq \epsilon$ for all $\hat{q} \in \bar{H}^h$.

Assuming agents truthfully report their signals, there are at least two hubs, and the identification conditions of Proposition 1 are satisfied, a designer can modify her identification strategy by using ϵ as denoted in Equation 2 to identify the characteristics of all objects with high probability. For example, suppose there are three possible characteristics of objects, q_1, q_2, q_3 and $\{q_1, q_2\} = s(q_2)$ and $q_3 = s(q_3)$. For identification purposes, the designer translates any report of q_2 or q_1 into q_2 . However, the Bayesian posterior assessment will inciate one of the options is more likely if $p(\hat{q} \in \{q_1, q_2\}|q_1) \neq p(\hat{q} \in \{q_1, q_2\}|q_2) \neq p(\hat{q} \in \{q_1, q_2\}|q_3)$. In other words, if the probability of an agent seeing seeing signal q_3 is different for all objects then the report $\hat{q} = q_2$ is still valuable evidence. If the number of agents (signals) increases, the designer will be able to pinpoint the characteristic of every object using the same techniques as in Proposition 1.

Corollary **1** : *If the conditions of Proposition 1 are satisfied for a sequence of markets $\mathcal{M}^1, \mathcal{M}^2, \ldots$ the probability of $\epsilon-$correctly identifying every object under the coarsened signal structure for ϵ satisfying Equation 2 converges to one when agents truthfully report their signals.*

4 Main Result

The previous section dealt with the identification power of a direct mechanism that incentivizes truthtelling. This section shows that it is possible to ensure truthtelling incentives by punishing each agent every time one of her reports is “incorrect,” and that the total payments made by any agent converges to zero in every realization of the market. In fact, even if every agent has to pay the maximum payment the mechanism can perscribe, the total payment still converges to zero. The punishment used resembles a Bayesian estimator with a 0/1 loss function; the designer solicits signals that maximize the posterior probability of being correct (hubs) and imposes a constant punishment whenever an agent is incorrect, regardless of the magnitude of the error.

One additional concern with equilibrium reporting are measuring the incentives of agents to truthfully report their signals. To begin, let $\beta_{m,m'} = p(q_{(n)} = q_m | \hat{q} = q_{m'})$, and let $m_{\max}(m') = \max_{m,m'} \beta_{m,m'}$. Let

$$\beta = \min_{m'} \left[m_{\max}(m') - \max_{m \neq m_{\max}(m')} \beta_{m,m'} \right]. \tag{3}$$

β can be though of as the opportunity cost for telling the truth, that is, it is an upper bound on how likely an agent will be to correctly identify an object by lying.

I now define mechanism $\bar{\varphi}$ as follows:

Step 1: Each agent i submits a report $\tilde{q}_{i,n}$ for each n.

Step 2: Let ϵ satisfy Equation (2) and say that $\bar{q}_{(n)} = q_m$ if n is $\epsilon-$identified as having characteristic q_m by assessment of agents. If n cannot be uniquely identified in this way, mark this object as “unidentifiable” and assign it a characteristic at random.

Step 3: Run mechanism φ using the characteristics $\bar{q}_{(N)}$ from *Step 2*.

Step 4: For each agent i let μ_i represent the number of objects for whom i's report $\tilde{q}_{i,n} \neq \bar{q}_{(n)}$. If an object is unidentifiable, augment μ_i by one, that is, count it as a disagreement.

Step 5: Charge each agent i a penalty of $\mu_i \cdot D \cdot \frac{2e^{-2I^k\delta^2}}{\beta-(\beta+1)2e^{-2I^k\delta^2}}$ where $\delta = \epsilon - \frac{1}{I^k}$.[11]

Recall that the goal for $\bar{\varphi}$ is important to create truthtelling incentives for all agents while minimizing waste. One criterion for evaluating the waste of this mechanism is to bound the worst case total sum of punishments, that is, what is the total waste when every agent has to pay the maximum penalty?

THEOREM **1** *: In a sequence of markets $\mathcal{M}^1, \mathcal{M}^2, \ldots$ with $N^k \in o\left(\frac{e^{2I^k\delta^2}}{I^k}\right)$ for all $\theta_\ell \in S$, as $k \to \infty$ mechanism $\bar{\varphi}$ is BNIC and the sum of penalties paid by all agents converges to zero in every realization.*

Note that the growth rate assumption is slightly more restrictive here than in Proposition 1, as there is an I^k in the denominator of this expression. The implication is that truthtelling and small

[11] In order to secure a sufficient punishment in every situation, the bound taken is quite loose. As a result, the denominator of the prescribed punishment may be negative or zero for small numbers of agents. $\bar{\varphi}$ can be modified to be constant in such situations and charge zero punishment for each agent.

punishments can be guaranteed in most cases, though not always, as when a designer with access to all signals can approximate an optimal matching.

Also, note that it is possible to designate one agent as a residual claimant. The market designer does not solicit reports from her, and transfers all collected punishments to this agent. By doing so, there is no waste as all punishments are just transfered to another agent.

I now discuss how to transform the monetary punishment in Theorem 1 into a non-monetary punishment that achieves similar incentive and non-wastefullness properties. The method for doing this is to create a mechanism, $\bar{\varphi}^{NM}$ that is identical to $\bar{\varphi}$ except that instead of punishing agents by charging a penalty it withholds agent i's consumption bundle probabilistically as a function of μ_i. In order to ensure that this method of punishment is successful, I assume two regularity conditions. First, a market is *consumption disjoint* if it is possible to separate the consumption bundles of agents and each agent cares only about what she consumes herself. Second, a sequence of markets and a mechanism satisfy *strict individual rationality* when agents, for any signals of characteristics they receive, will always achieve greater expected utility under any feasible allocation than they would by consuming nothing, and that this difference is bounded away from zero.

DEFINITION **2** *: $\mathcal{M}$ is* consumption disjoint *if* $\forall q_{(N)}$, $\forall i \in I$, $\forall I^{'} \subset I \setminus \{i\}$, $\forall b \in B$ *with* $b_i \neq \emptyset_i$*:* $\exists b' \in B$ *such that* $b_i = b_i^{'}$ *and* $b_j = \emptyset\ \forall j \in I^{'}$*, and* $u_i(b, q_{(N)}) = u_i(b', q_{(N)})$*.*

Matching models frequently exhibit this property, as each agent cares only about her own partner. On the other hand, markets for the provision of public goods are not consumption disjoint.

DEFINITION **3** *: Let $\emptyset_i$ be agent i's outside option. A sequence of markets $\mathcal{M}^1, \mathcal{M}^2, ...$ and mechanism φ exhibit* strict individual rationality *if* $\exists \psi > 0$ *such that* $\inf_k \mathbb{E}\left[u_i(\varphi(\mathcal{M}^k), q_{(N^K)}) - u_i(\emptyset_i, q_{(N^K)}) | \hat{q}_{i,(N^K)}\right] \geq \psi$*.*

With these regularity conditions, I define mechanism $\bar{\varphi}^{NM}$ by replacing *Step 5.* in mechanism $\bar{\varphi}$ with:

Step 5': With probability $\frac{1}{\psi}\mu_i \cdot D \cdot \frac{2e^{-2I^k\delta^2}}{\beta-(\beta+1)2e^{-2I^k\delta^2}}$ replace b_i with $\emptyset_i$.

Mechanism $\bar{\varphi}^{NM}$ (under the assumed regularity conditions) punishes agents by cancelling agents' allocations with small probability ex-post.

COROLLARY **2** *: If a sequence of markets $\mathcal{M}^1, \mathcal{M}^2, ...$ and mechanism φ are consumption disjoint, exhibit strict individual rationality, and $N^k \in o\left(\frac{e^{2I^k\delta^2}}{I^k}\right)$ then as $k \to \infty$ mechanism $\bar{\varphi}^{NM}$ is BNIC and the number of agents whose allocations are cancelled converges in probability to 0.*

5 Additional Results

5.1 Costly signal acquisition

Thus far, the analysis has assumed that agents exogenously receive signals. More realistically, agents will have to invest time or effort to recieve signals of object characteristics. As the analysis in this paper hinges on many agents revealing signals of many objects, a major concern is that identifying the characteristics of all objects will be prohibitively costly. In this section I explore this by allowing agents to endogenously pick their signal strength, with more informative signals being more costly.

In order to evaluate T objects with identification strength $\epsilon \in [0, a]$ where $a < 1$, I assume an agent must pay $T \cdot c(\epsilon)$, where $c(0) = 0$ and $c(\cdot)$ is strictly increasing and the d^{th} derivative $c^{(d)}(0)$ exists for all d. To identify the characteristics of all objects with asymptotically 0 cost it must be the case that $T^k \cdot c(\epsilon^k) \to 0$. However, when I^k and N^k grow at similar rates, it must be that $T^k \to \infty$ as otherwise it would be impossible to have an arbitrarily large number of signals for each object, a necessary condition to correctly identify the characteristics of all objects. Therefore, it must also be the case that $c(\epsilon^k) \to 0$ sufficiently quickly.

An important part of equilibrium analysis relies on relating ϵ, which measures the identification power of the market designer, and β, which measures incentives of agents to truthfully report signals. In previous sections both of these were constants and were washed away in the limit as the market size grew. It is easy to see that weakening the signal by reducing ϵ can reduce β as agents are more likely to believe their priors. When this is the case, the relative speed at which both ϵ^k and β^k converge to zero matters. Furthermore, as ϵ^k skrinks identification may become impossible as the set of hubs may shrink to a singleton. To put some structure on the problem and simplify the analysis I will assume that there are two possible object characteristics, each of which is equally likely ex-ante, and that for any $\epsilon > 0$ signals are modal. It is possible to extend the analysis and proof structure to allow for more possible object characteristics under additional symmetry assumptions that bound the ratio of ϵ^k to β^k for large k.

As I show, when identification is possible with asymptotically zero cost, it is achieved by having all agents report (weak) signals of all objects.[12] I give conditions under which a sequence of signals strengths and punishments (for incorrect reports compared to the wisdom of the crowds, as before), $\left\{\{\epsilon_i^k, p_i^k\}_{i=1,\ldots,I^K}\right\}_{k=1,2,\ldots}$, exists such that all agents follow the plan in equilibrium, truthfully report all signals, all characteristics are asymptotically identified, and total waste goes to 0.

THEOREM **2** *: Let $N^k \sim \left(I^k\right)^\alpha$ for some $\alpha \geq 0$, and suppose $Q = \{q_1, q_2\}$, each of which is equally likely ex-ante, and that for any $\epsilon > 0$, $p(\hat{q}_{i,n} = q_{(n)}) > \frac{1}{2}$. Then there exists a sequence $\left\{\{\epsilon_i^k, p_i^k\}_{i=1,\ldots,I^K}\right\}_{k=1,2,\ldots}$ such that as $k \to \infty$ all agents report signals on all objects and*

1. *if the designer can pick ϵ_i^k for all agents and ensure truthful reporting, all objects are correctly identified with zero total cost if and only if $c^{(d)}(0) = 0$ for all $d \leq 2\alpha + 2$,*

2. *If the designer can pick ϵ_i^k for all agents and ensure truthful reporting, all objects are correctly identified with zero per capita cost if and only if $c^{(d)}(0) = 0$ for all $d \leq 2\alpha$,*

3. *the reporting game specified by the mechanism is BNIC, and in the truthtelling equilibrium all objects are correctly identified, total cost is zero, and all agents pick the specified ϵ_i^k if and only if $c^{(d)}(0) = 0$ for all $d \leq 2\alpha + 3$, and*

[12] Bohren and Kravitz (2016) study a principal who crowdsources the identification of unverifiable information to multiple agents. They similarly find that incentive constraints can be relaxed if agents are called upon to assess many objects.

4. the reporting game specified by the mechanism is BNIC, and in the truthtelling equilibrium all objects are correctly identified, per capita cost is zero, and all agents pick the specified ϵ_i^k *if and only if* $c^{(d)}(0) = 0$ *for all* $d \leq 2\alpha + 1$.

Considering $\alpha \geq 0$ is sufficient to describe all interesting relative growth rates. $\alpha = 0$ means that the number of objects in the market is constant in the number of agents, and one can trace out everything slower than polynomial growth with small α as the conditions given in Theorem 2 are the same for $\alpha = 0$ and α sufficiently close to zero. Also, any growth rate faster than polynomial growth, i.e. $\left(I^k\right)^\alpha \in o(N^k)$ for all $\alpha > 0$, such as exponential growth, implies that a necessary condition to identify all objects at asymptotically zero cost is that $c(\gamma) = 0$ for some $\gamma > 0$, which means some information is free.

Consider specifically the case in which N^k is constant ($\alpha = 0$). Theorem 6 of Martinelli (2006) gives the same condition, $c(0) = c^{'}(0) = c^{''}(0) = c^{'''}(0) = 0$, for a Condorcet Jury model with voluntary information acquisition and a particular family of agent preferences to make the "correct decision" at asymptotically zero aggregate cost. In such a jury model, not all agents participate. One interpretation of the results of Theorem 2 is that with arbitrary utility functions, all agents will gather small amounts of information (but enough to collectively determine the "correct decision") and participate in the vote if there is some social cost of non-participation.

Another interpretation of this theorem is the cost a principal has to pay experts to gather and report costly signals. Experts have to be compensated for their effort in gathering signals, and also have to be incentivized to tell the truth (in order to ensure that the offered contract is individually rational to experts, p^k can be a reward for agreeing with the market consensus instead of a punishment for disagreeing). Gerardi and Yariv (2007) consider this problem, although they assume that there is only one object with an unknown quality, and the principal hires at most two experts from a large pool. The case in which effort and signals are public corresponds to the market designer result of Theorem 2. Theorem 2 can be used to study conditions under which it is optimal for a principal to contract out the information gathering to many agents. One interesting finding of this theorem is that fewer restrictions are necessary on the cost function to satisfy agents' incentive constraints, both in terms of picking the correct signal strength and truthfully reporting signals, than to move from zero per capita waste to zero total waste.

5.2 Unknown Utility Functions

One assumption carried throughout the paper is that the utility functions of all agents are known. Although this may be true in some circumstances, many markets feature agents with heterogeneous and unknown preferences. The analysis above gives conditions under which a designer can, with high probability, ascertain the unknown characteristics of all objects, nevertheless, the posterior belief of each object's characteristic may differ, as mechanism $\bar{\varphi}$ rounds the posterior assessment of the characteristic of each object to the nearest discrete characteristic level. Even if a market designer creates a mechanism that gives incentives for truthtelling when characteristics are known, these incentives will generally not carry through when characteristics are merely known with high probability.

I investigate the recovery of incentives to truthfully reveal preferences in two contexts. First, suppose that the characteristic of every object is known by an agent. For example, a producer may know the characteristic of her own product whereas her peers only observe signals of its characteristic, or a worker may know her own characteristic whereas interviewing firms observe a signal of her characteristic.

Definition **4** *: φ is* strategy-proof *if the revelation game induced by φ has an equilibrium in weakly dominant strategy for each agent i to truthfully report her utility function.*

When one agent has perfect knowledge of an object, a designer can create a "check and balance"–she can run mechanism $\bar{\varphi}$ with additional steps.

> *Step 2.1:* Solicit reports of u_i from each agent i.
> *Step 2.2:* Solicit agent $n^{'}$ (who knows the characteristic of item n) to report on the characteristic of item n, $\bar{s}_n$. If $\bar{q}_{(n)} = \bar{s}_n$ then continue as before. Otherwise, charge agent $n^{'}$ a penality of $\left(1 + 2e^{-2I^k\delta^2}\right) \cdot D$ and define $\bar{q}_{(n)} \equiv \bar{s}_n$.

Call the mechanism with these two steps added $\bar{\varphi}^{CB}$ By adding these extra steps, in a truth-telling equilibrium, agent $n^{'}$ with full knowledge of $q_{(n)}$ will have an incentive to report $\bar{s}_n$ truthfully.

Proposition **2** *: Suppose for each object n there is an agent $n^{'}$ who knows q_n. If a sequence of markets $\mathcal{M}^1, \mathcal{M}^2, ...$ satisfies $N^k \in o\left(\frac{e^{2I^k\delta^2}}{I^k}\right)$ then as $k \to \infty$ for any strategy-proof mechanism φ, the game induced by $\bar{\varphi}^{CB}$ has a Bayes Nash equilibrium in undominated strategies in which all agents truthfully report signals and utility functions and expected total payments converge to 0.*

The preceeding mechanism $\bar{\varphi}^{CB}$ uses a check-and-balance system to incentivize all parties to truthfully report the characteristics of each object and their own utility functions. The premise of this mechanism makes the assumption that at least one person knows the characteristic of each object with probability 1. I show that a similar claim holds when no agent knows the characteristic of a particular object in an *ordinal mechanism.*

Definition **5** *: Let $\mathcal{M}$ and $\mathcal{L}$ be two markets which differ only in the utility functions of agents, $\{u_i\}_{i\in I}$ and $\left\{u_i^{'}\right\}_{i\in I}$, respectively. φ is an* ordinal mechanism *if $\varphi(\mathcal{M}) = \varphi(\mathcal{L})$ whenever $u_i(b, N_q) \geq u_i(\tilde{b}, \tilde{N}_q)$ if and only if $u_i^{'}(b, N_q) \geq u_i^{'}(\tilde{b}, \tilde{N}_q)$ for all $i, b, \tilde{b}, N_q, \tilde{N}_q$.*

Ordinal mechanisms operate independently of the cardinal values of preferences. Examples of such mechanisms abound in the matching literature, including Deferred Acceptance, Top Trading Cycles and Serial Dictatoriship mechanisms.

Definition **6** *: Utility functions $\{u_i\}_{i\in I}$ are* bounded-repsonsive *if $u_i(b, N_q) = \sum_{n\in N} b_{i,n} \cdot v(x_{i,n}, n_q)$ for all feasible b_i where $x_{i,n} \overset{i.i.d.}{\sim} F_i$, F_i has full support,dF_i is uniformly bounded above by $g > 0$ for all i, and v_i is continuous in its first argument.*

This is a generalization of a concept known as responsive preferences found in the matching literature (Roth (1985)).

I now define mechanism $\bar{\varphi}^O$ as $\bar{\varphi}$ with the following additional step.

Step 2.1: Solicit reports of u_i from each agent i.

PROPOSITION **3** *: Suppose that φ is ordinal and strategy-proof and utility functions $\{u_i\}_{i \in I^k}$ are bounded-repsonsive. Then in a sequence of markets $\mathcal{M}^1, \mathcal{M}^2, ...$ with $\left(N^k\right)^2 \in o\left(\frac{e^{2I^k \delta^2}}{I^k}\right)$ as $k \to \infty$ the game induced by $\bar{\varphi}^O$ has a Bayes Nash equilibrium in undominated strategies in which all agents truthfully report signals and additionally, utility functions with probability approaching 1 and total payments converge to 0.*

Notice that the growth rate condition is more restrictive (the condition is on $\left(N^k\right)^2$ not N^k) and the growth rate condition is necessary for the truthful reporting of utility functions, whereas in Proposition 2 the growth rate condition was only applied to the claim about the sum of punishments. The reason is that mechanism $\bar{\varphi}^O$ induces truthtelling when the private cardinal values for all objects are sufficiently far apart, which, under the independence assumption, is harder to maintain as the number of objects increases.

5.3 Coordinated lying

The analysis thus far has been concerned with designing a mechanism such that no agent has any incentives to unilaterally lie. Nevertheless, mechanism $\bar{\varphi}$ may be succeptible to coordinated lies by a group of agents. For example, there can even exist an equilibrium in which all agents report the "opposite" of their signals. This section shows that it is possible for a modification of mechanism $\bar{\varphi}$ to remove incentives for coordinated lying. Expositionally, I present this as a strategic market designer playing a strategy to block coalitions of agents from misreporting signals.

In particular, suppose that the designer has the option to suggest to each agent i to correctly reveal her signals, i.e. $S_i \in \{0, 1\}$. If $S_i = 1$ then with independent probability $\eta < 1$ agent i will truthfully reveal all of her signals, and if $S_i = 0$ then agent i will remain strategic. In other words, each agent has an η probability of being a naive type and trusting the recommendation of the market designer. The designer can therefore, with high probability, guarantee a nearly known proportion of truthful reports as the market grows large, but she does not know which agents are following her suggestion, which are strategically lying, and which are strategically being truthful.

Nevertheless, by creating aggregate uncertainty about the proportion of agents who are non-strategic, tightening the ϵ bounds used for identification and increasing the punishment when objects fail to be identified, the designer can ensure that no coalitions will form to jointly misreport signals.

Taking the designer as a player whose strategy space consists of a sequence $S_1, ..., S_I$, the designer can create aggregate uncertainty in the proportion of non-strategic players by mixing between S_is in a correlated manner. In particular, suppose the designer mixes between three alternatives with equal probability: $S_i = 1$ for all i, $S_i = 0$ for all i and $S_i = 1$ independently with probability ρ for each i. [13] Upon seeing $S_i = 1$ ($S_i = 0$) agent i will attribute a Bayes posterior of $\frac{1}{1+\rho}$ $\left(\frac{1}{2-\rho}\right)$ of the designer having chosen strategy $S_i = 1$ for all i ($S_i = 0$ for all i). Meanwhile, if the designer (somewhat carefully) chooses a sequence of $\epsilon^k \to 0$ then any coalitional lie will, with high probability, leave the

[13] Note that the analysis of this section is unchanged if, instead of treating the market designer as a player, the parameter determining the probability of an agent being naive is drawn uniformly from $\{0, \rho \cdot \eta, \eta\}$.

object in question unidentified for at least one of the actions chosen by the designer.[14] Therefore, by ammending mechanism $\bar{\varphi}$ to punish all agents large amounts in the unlikely case (in equilibrium) that an object is unidentified, no coalition will have an incentive to lie.

PROPOSITION **4** *: In a sequence of markets $\mathcal{M}^1, \mathcal{M}^2, ...$ with $\left(N^k\right)^2 \in o\left(\frac{e^{2I^k\delta^2}}{I^k}\right)$ and $\eta > 0$ probability of each agent being naive, as $k \to \infty$ there exists a BNIC mechanism that identifies all object characteristics, total expected payments converge to 0 in the truthtelling equilibrium, and no coalition of agents benefits from jointly lying.*

6 Robustness and Conclusion

This paper studies how crowdsourcing can aggregate information in markets in order to ensure near-optimal outcomes. In settings of interdependent preferences, the wisdom of the crowds is sufficient to identify the characteristics of all objects with high probability as the market grows, even when the number of objects to estimate grows significantly faster than the number of agents. However, agents may not have incentives to report these signals truthfully as they may wish to game the system. By introducing small penalties for reports that do not align with the collective opinion of other agents, either by imposing a monetary transfer or cancelling an agent's allocation ex-post, a mechanism designer can remove the incentive of agents to game the system. I show that the sum of punishments across all agents necessary to restore good incentives converges to zero. Therefore, a market designer can use crowdsourced information to approximate optimal allocations.

The results of this paper are applicable to many economically interesting markets such as labor market matching, competitive equilibrium, and peer assessment. The mechanism I propose is independent on the specifics of the market in question. The rates of convergence discussed in the formal results are exponential, meaning that a "large market" need not be prohibitively large. As discussed in Section 2, reasonably-sized markets have sufficiently many agents that the techniques of this paper will ensure proper incentives and approximate desirable allocations with minimal waste.

One observation is that the generality of the setting of this paper leads to quite liberal bounds on identification and punishment, but this also imples that results are relatively "detail-free" and robust to lack of full information on distributions on the part of agents in the model (Wilson (1987)). The mechanism presented depends on two values to reprsent the strength of agents' signals, ϵ and β. But these values can be represented as bounds as opposed to exact values. For example, suppose the only knowledge a designer has is that $p\left(q_{(n)} = q_\ell | \hat{q}_{i,n} = q_\ell\right) \geq \Omega > \frac{1}{2}$ for all i and ℓ, that is the probability each agent places on the event that the quality of an item is equal to the signal she observes is bounded away from $\frac{1}{2}$. Then the designer can utilitze the mechanism in this paper by setting $\beta = 2\Omega$ and $\epsilon = \Omega$, and all of the properties of the mechanism will hold.[15] A recent mechanism proposed by McLean and Postlewaite (2018) that is specificially designed to be robust to various information settings (indeed, it is referred to by the authors as "a very robust auction mechanism) makes a nearly-identical assumption on distributions. To my knowledge, no mechanism

[14] Recall that an object is unidentified if the empirical distribution of signals regarding its characteristic is not within distance ϵ of the theoretical distribution for any characteristic with respect to the supnorm.

[15] This observation is the basis for the back-of-the-envelope calculations presented in Section 2.

in the literature that incentivizes truthtelling with vanishingly small waste can operate on fewer details than the mechanism in this paper, and that in McLean and Postlewaite (2018).

In a way, this paper also gives a theoretical justification for the pervasive assumption of full-information in the market design literature–allocations derived withthis assumption can be approximated without it. Since the full-information assumption buys significant tractibility, it is a logical assumption to make. Even in some more complicated settings, such as when signals are costly to observe or agents have private information about their preferences, the results of this paper show that the driving force of the wisdom of the crowds is powerful enough to approximate optimal allocations. Of course, this may not hold in certain markets in which signals are (perfectly) correlated across agents or when it is prohibitively costly for agents to receive non-trivial signals for many objects. The study of such markets presents new challenges and is left for further study.

References

Mohammad Akbarpour and Afshin Nikzad. Approximate random allocation mechanisms. *Mimeo*, 2017. 4

William H. Batchelder and A. Kimball Romney. Test theory without an answer key. *Psychometrika*, 53(1):71–92, 1988. 1

Sushil Bikhchandani. Stability with one-sided incomplete information. *Journal of Economic Theory*, 168:372–399, March 2017. 1

Aislinn Bohren and Troy Kravitz. Optimal contracting with costly state verification, with an application to crowdsourcing. *Mimeo*, 2016. 12

Archishman Chakraborty, Alessandro Citanna, and Michael Ostrovsky. Two-sided matching with interdependent values. *Journal of Economic Theory*, 145(1):85–105, January 2010. 1, 8

Yeon-Koo Che, Jinwoo Kim, and Fuhito Kojima. Efficient assignment with interdependent values. *Journal of Economic Theory*, Forthcoming, 2015. 1

Brian Coulter and Joel D. Shapiro. A mechanism for libor. *mimeo*, 2015. 2

Vincent P. Crawford. Comparative statics in matching markets. *Journal of Economic Theory*, 54: 389–400, 1991. 1

Jacques Cremer and Richard P. McLean. Optimal selling strategies under uncertainty for a monopolist when demands are interdependent. *Econometrica*, 53(2):345–361, March 1985. 1

Jacques Cremer and Richard P. McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56(6):1247–1257, November 1988.

Yuji Fujinaka and Toshiji Miyakawa. A model of housing markets with interdependent values. *mimeo*, 2015. 1

Patricia Funk. Theory and evidence on the role of social norms in voting. *mimeo*, 2005. 7

David Gale and Lloyd S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, January 1962. 2

Dino Gerardi and Leeat Yariv. Costly expertise. *Games and Economic Behavior*, 62:436–459, 2007. 5.1

Dino Gerardi, Richard P. McLean, and Andrew Postlewaite. Aggregation of expert opinions. *Games and Economic Behavior*, 65(2):339–371, 2008. 1

Tadashi Hashimoto. The generalized random priority mechanism with budgets. *Journal of Economic Theory*, 177:708–733, September 2018. 4

Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. *In Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 53–62, January 2005. 1

Philippe Jehiel and Benny Moldovanu. Efficient design with interdependent valuations. *Econometrica*, 69(5):1237–1259, September 2001. 1

Philippe Jehiel, Moritz Meyer-ter Vehn, Benny Moldovanu, and William R. Zame. The limits of ex post implementation. *Econometrica*, 74(3):585–610, 2006. 1

Fuhito Kojima and Parag A. Pathak. Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review*, 99(3):608–627, 2009. 1

Harold A. Linstone and Murray Turoff. *The Dephi Method: Techniques and Applications*. Addison-Wesley, 2002. 1

Qingmin Liu, George Mailath, Andrew Postlewaite, and Larry Samuelson. Stable matching with incomplete information. *Econometrica*, 82(2):541–587, March 2014. 1, 8

Cesar Martinelli. Would rational voters acquire costly information? *Journal of Economic Theory*, 129:225–251, 2006. 1, 5.1

Maggie McKee. Mathematical "truth serum" promotes honesty. *NewScientist*, 2014. 3

Richard P. McLean and Andrew Postlewaite. Informational size and incentive compatibility. *Econometrica*, 70(6):2421–2453, November 2002. 1, 3.3

Richard P. McLean and Andrew Postlewaite. Informational size and efficient auctions. *Review of Economic Studies*, 71:809–827, June 2003. 5, 3.3

Richard P. McLean and Andrew Postlewaite. Implementation with interdependent valuations. *Theoretical Economics*, 10:923–952, 2015.

Richard P. McLean and Andrew Postlewaite. A dynamic non-direct implementation mechanism for interdependent value problems. *Games and Economic Behavior*, 101:34–48, 2017.

Richard P. McLean and Andrew Postlewaite. A very robust auction mechanism. *mimeo*, 2018. 6

Nima Mousavi. How tight is chernoff bound? *https://ece.uwaterloo.ca/ nmousavi/Papers/Chernoff-Tightness.pdf*, 2010.

Chris Piech, Jonathan Huang, Zhenghao Chen, Chuong Do, Andrew Ng, and Daphne Koller. Tuned models of peer assessment in moocs. *mimeo*, 2013. 2

Drazen Prelec. A bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004. 1, 3

Natalia Rigol and Benjamin Roth. Paying for the truth: The efficacy of a peer prediction mechanism in the field. *mimeo*, 2016. 3

Alvin E. Roth. The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36:277–288, 1985. 5.2

Alvin E. Roth and Elliott Peranson. The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89(4):748–780, 1999. 1

Ray Weaver and Drazen Prelec. Creating truth-telling incentives with the bayesian truth serum. *Journal of Marketing Research*, 50(3):289–302, 2013. 3

Robert Wilson. Game theoretic analysis of trading processes. In T. Bewley, editor, *Advances in Economic Theory*. Cambridge University Press, 1987. 6

Appendix

A Failure of Information Aggregation: Examples

EXAMPLE **3** *: There are I^k and N^k of an object with $I^k > N^k$ for all k. Each object is differentiated only by quality, with each seller i equally likely to create a high quality object x_i or a low quality object y_i, so that seller i's endowment $e_i \in \{(1,0),(0,1)\}$ each with equal probability.*

Each buyer wishes to buy at most one object, and only wishes to buy a high quality object, at the lowest possible price. Formally, the utility of each buyer n is

$$u^n(x,y,t) = \left[\sum_{j=1}^{I^k} x_j^n - y_j^n\right] \cdot \mathbf{1}_{\left\{\sum_{j=1}^{I^k} x_j^n + y_j^n \leq 1\right\}} - \sum_{j=1}^{I^k} t_j^n$$

where t_j^n is the payment the buyer makes to seller j. Sellers value only money, and each seller wishes to sell her product for the highest price,

$$u_i(x,y,t) = \sum_{m=1}^{N^k} t_i^m.$$

Suppose qualities are common knowledge and there are H^k objects of high quality. In a Walrasian Equilibrium, if $H^k > N^k$ then all high quality objects are free and all buyers receive exactly one object, if $H^k = N^k$ then all high quality objects are priced at some $t^* \in [0,1]$ and each buyer buys exactly one object, if $H^k < N^k$ then the price of each object is 1 and all sellers of high quality objects sell their object. In all cases, the low quality objects are free and unclaimed.

Now suppose that qualities are unknown. Each seller receives a signal of the quality of each object, while buyers cannot differentiate between objects of different qualities. A market designer solicits reports of qualities and computes a Walrasian Equilibrium based on these reports. Note that the conditions of Proposition 1 are satisfied as $I^k \to \infty$, and so the qualities of all objects will be ascertained asymptotically if the signals reported are truthful. Let $\bar{H}^k$ denote the market assessment for the number of high quality objects based on the reports.

Now consider the incentives for seller i. I claim that seller i has a weakly dominant strategy to report that all other sellers have low quality objects and that she has a high quality object. Since i's probability of selling her object and the price at which she sells it are non-increasing in $\bar{H}^k$, she does no worse (and possibly strictly better) by reporting that all objects are low quality. Since all sellers have the same incentives, all objects will be judged to be of low quality.

EXAMPLE **4** *: There are I laws of either high or low quality. A non-partisan and unbiased president wishes to enact high quality laws. She relies on the advice of the senate to determine which policies she should sign into law (she can veto bills passed by the senate, or enact failed legislation through executive orders). Each senator is partially informed, receiving a signal $p > \frac{1}{2}$ of the quality of each law. However, senators are partisan, with senators from the majority party wishing to enact even-numbered laws of either quality, and not enacting odd-numbered laws of either policy. Minority party senators have opposite preferences.*

Therefore, each senator has a weakly dominant strategy to vote for a law if and only if it conforms with her party platform. By following the recommendation of the senate, the president will enact $\lfloor \frac{n}{2} \rfloor$ laws. Note that if the prior probability of each law being high quality is greater than $\frac{1}{2}$ and the president's utility from enacting a high quality law minus the utility from enacting a low quality law is greater than 0, then she would be better off ignoring the senate and enacting every law.

B Proofs

Proof of Proposition 1:

Proof of Part 1.:

The mathematical basis for the proof is based on the following result.

LEMMA **1** [Dvoretzky, Kiefer, Wolfowitz (1956) and Massart (1990)]:

For any $\lambda > 0$, $p\left(\sqrt{I^k}\sup_x|\hat{F}_{I^k}(x) - F(x)| > \lambda\right) \leq 2e^{-2\lambda^2}$.

Rearranging the DKW inequality, then

$$p\left(\sup_{\hat{q}}\left|\hat{F}_{I^k}\left(\hat{q}|q_{(n)}\right)-F\left(\hat{q}|q_{(n)}\right)\right|>\frac{\lambda}{\sqrt{I^k}}\right)\leq 2e^{-2\lambda^2}.$$

Let $\lambda=\epsilon\sqrt{I^k}$. Then the DKW inequality further implies that

$$p\left(\sup_{\hat{q}}\left|\hat{F}_{I^k}\left(\hat{q}|q_{(n)}\right)-F\left(\hat{q}|q_{(n)}\right)\right|>\epsilon\right)\leq 2e^{-2I^k\epsilon^2}$$

so the probability of agents incorrectly identifying the characteristic of n is declining exponentially in the number of agents. However, we are interested in the probability that the characteristics of *all* objects are correctly identified. The probability of incorrectly identifying even a single $n\in N^k$ is bounded above by

$$N^k\cdot 2e^{-2I^k\epsilon^2}\equiv\chi_k \tag{4}$$

Then $\chi_k\to 0$ if and only if $2\epsilon^2I^k-\log N^k\to\infty$, which occurs if and only if $\forall\gamma>0\,\exists K:\ \forall k>K\ I^k\geq\frac{1}{2\epsilon^2}\log N^k+\gamma$. This establishes the "if" direction.

Proof of Part 3.:

To establish the converse, consider the following result:

LEMMA **2** [Mousavi (2010)]: *Let $X_1,...X_{I^k}$ be independent, Bernoulli random variables with success probability $\rho\leq\frac{1}{4}$. For any $t>0$, $p\left[\sum_{j=1}^{I^k}X_j-I^k\rho>t\right]\geq\frac{1}{4}e^{\frac{-2t^2}{I^k\rho}}$.*

Dividing through the first term of the above inequality, we see that

$$p\left[\frac{1}{I^k}\sum_{j=1}^{I^k}X_j-\rho>\frac{t}{I^k}\right]\geq\frac{1}{4}e^{\frac{-2t^2}{I^k\rho}}$$

Taking $t=\epsilon I^k$, we can rewrite this as

$$p\left[\frac{1}{I^k}\sum_{j=1}^{I^k}X_j-\rho>\epsilon\right]\geq\frac{1}{4}e^{\frac{-2\epsilon^2I^k}{\rho}}$$

As $|Q|>3$ there must be some q_o for any object having characteristic $q_{(n)}$ such that $\rho=p\left(\hat{q}=q_o|q_{(n)}\right)\leq\frac{1}{4}$, i.e. the probability of getting signal q_o given the object's characteristic is $q_{(n)}$. Then $\frac{1}{I^k}\sum_{j=1}^{I^k}X_j$ is the empirical distribution of getting signal q_o, and the above inequality gives a lower bound on the probability of the empirical distribution being farther than ϵ from the expected distribution at q_o. As $\epsilon-$identification requires the empirical distribution is no farther than ϵ *over the full support* of Q, $\frac{1}{4}e^{\frac{-2\epsilon^2I^k}{\rho}}$ is a lower bound on the probability of incorrectly identifying the object characteristic. Note that both this value and the expoential term used in the "if" direction of the proof are both on the order of e^{-I^k}. Therefore, reworking Equation (4) with the lower bound gives the same requirement on the relative growth rates of $N^k$and I^k, up to a constant factor.

Proof of Part 2.:

Follows from Equation (4).

■

Proof of Corollary 1:

Assume the hypothesis of Proposition 1. For any n suppose agents can distinguish between two characteristics, that is, there exists $q_o$and q_p such that $F(q_o) \neq F(q_p)$. Now consider partion $\bar{\mathcal{F}}$. Recall that$\bar{\mathcal{F}}(q_o) = \bar{\mathcal{F}}(q_p)$ only if $p\left(\hat{q} \in \bar{H}^h(q)|q_o\right) = p\left(\hat{q} \in \bar{H}^h(q)|q_p\right)$ for all $\bar{H}^h(q)$. Since there are at least two hubs and signals have full support, this occurs only when

$$\sum_{\hat{q} \in \bar{H}^h(q)} p\left(\hat{q}|, q_o\right) = \sum_{\hat{q} \in \bar{H}^h(q)} p\left(\hat{q}|q_p\right) \text{ for all } q \in Q. \tag{5}$$

Since by assumption $\bar{H}^h(q) \neq \bar{H}^h(q')$ for some $q, q' \in Q$, $\bar{\mathcal{F}}(q_o) = \bar{\mathcal{F}}(q_p)$ for a non-generic set of distributions (i.e. the set of distributions for which (5) does not hold is open and dense). The remainder of the proof follows from reapplying the logic of the proof of Proposition 1.

■

Proof of Theorem 1:

I first prove that the proposed punishment yields a truthtelling equilibrium, and second show that the total punishment converges to zero.

1. Suppose that an agent i of recieves signal $q_{m'}$ from object n. From Proposition 1 (and Corollary 1) the probability that i can change market assessment of n is bounded above by $2e^{-2I^k\delta^2}$. By assumption, the most i can gain from changing the market assessment (before taking the punishment into account) is D utils. I now give (upper and lower, respectively) bounds on the probability that i will have to pay a penalty, when the penalty is assessed when i's report differs from the market assessment for n. An upper bound on having the pay the punishment from telling the truth is
$$1 - m_{\max}(m')\left(1 - 2e^{-2I^k\delta^2}\right) \tag{6}$$
where (6) represents the case that i's signal is correct and the market assessment is also correct. On the other hand, a lower bound on having to pay the punishment by lying is
$$\left(1 - \max_{m \neq m_{\max}(m')} \beta_{m,m'}\right)\left(1 - 2e^{-2I^k\delta^2}\right) \tag{7}$$

where (7) represents the case that i reports the second most likely characteristic given her signal, is incorrect, and the market correctly identifies the characteristic of n. Combining (6) and (7) together with the largest possible benefit of changing the market assessment, punishment p^k is sufficient if

$$-\left(1 - m_{\max}(m')\left(1 - 2e^{-2I^k\delta^2}\right)\right)p^k \geq 2De^{-2I^k\delta^2} - \left(1 - \max_{m \neq m_{\max}(m')} \beta_{m,m'}\right)\left(1 - 2e^{-2I^k\delta^2}\right)p^k \tag{8}$$

Solving for the smallest p^k that satisfies (8) for any $q_{m'}$ yields

$$p^k = D \cdot \frac{2e^{-2I^k\delta^2}}{\beta - (\beta+1)2e^{-2I^k\delta^2}} \tag{9}$$

Noting that an agent could potentially gain D utils by changing the market assessment for even a single object yields the desired result.

2. To show that the sum of punishments converges to zero, note that the worst case total punishment is

$$I^k \cdot N^k \cdot D \cdot \frac{2e^{-2I^k\delta^2}}{\beta - (\beta+1)2e^{-2I^k\delta^2}} \tag{10}$$

which corresponds to the case in which every agent incorrectly identifies the characteristics of every object (or, alternatively, that every object is misidentified). However, with the ongoing assumption that $N^k \in o\left(\frac{e^{2I^k\delta^2}}{I^k}\right)$ it becomes clear that (10) converges to zero as $k \to \infty$.

■

Proof of Theorem 2:

In what follows, I will use write$\epsilon^k = \min_{i \in I^k} \epsilon_i^k$ (and $\beta^k = \min_{i \in I^k} \beta_i^k$), the smallest ϵ (β) chosen in market k. Note that the market designer can simply "add noise" to each agent's reports to lower the signal strength equal to ϵ^k to preserve the identification methods discussed in this paper. As I will show below, this is mostly a notational point, as the difference in the chosen ϵ_i^k between agents converges to zero very quickly.

I now prove a lemma that is useful in relating β and ϵ.

LEMMA **3** *: As* $\epsilon^k \to 0$, $\frac{\beta^k}{\epsilon^k} \to 4$.

Under the assumption of two object types, equal priors and signal modality, it is easy to see that

$$\beta^k =$$

$$p(q_{(n)} = q_1|\hat{q}_{i,n} = q_1) - p(q_{(n)} = q_2|\hat{q}_{i,n} = q_1) =$$

$$\frac{p(\hat{q}_{i,n} = q_1|q_{(n)} = q_1) - p(\hat{q}_{i,n} = q_2|q_{(n)} = q_1)}{p(\hat{q}_{i,n} = q_1)} \to$$

$$\frac{p(\hat{q}_{i,n} = q_1|q_{(n)} = q_1) - p(\hat{q}_{i,n} = q_2|q_{(n)} = q_1)}{\frac{1}{2}}$$

where the second equality comes from Bayes rule and the convergence comes from the assumption of signal modality and $\epsilon^k \to 0$. Furthermore,

$$\epsilon^k = \frac{p(\hat{q}_{i,n} = q_1|q_{(n)} = q_1) - p(\hat{q}_{i,n} = q_2|q_{(n)} = q_1)}{2}$$

Combining these two equations gives the desired result.

■

Returning to the proof of the theorem, I show that the desired properties yield constraints on $c(\cdot), \beta^k, \epsilon^k$ and p^k and the intersection of these constraints is non-empty if and only if the respective derivative condition is satisfied. Throughout I use that $\left(I^k\right)^\alpha \sim N^k$. Note that the assumption that all agents are instructed to acquire information on all objects is without loss of generality, as replacing I^k with $J^k < I^k$ does not easy any of the constraints in the limit.

Proof of Point 3:

First, the restriction that all characteristics are correctly identified requires that

$$I^k - \frac{1}{2\left(\epsilon^k\right)^2} \log N^k \to \infty. \tag{11}$$

Second, the restriction that the total cost paid in information aggregation goes to zero requires that

$$N^k I^k \left(p^k + c(\epsilon^k)\right) \to 0. \tag{12}$$

While the alternative that per capita costs go to zero requires that

$$N^k \left(p^k + c(\epsilon^k)\right) \to 0 \tag{13}$$

Third, the restriction that all agents truthfully report signals is implied by the condition of Theorem 1, that is,

$$p^k \geq \frac{2e^{-2I^k\left(\delta^k\right)^2}}{2\epsilon^k - (2\epsilon^k + 1)2e^{-2I^k\left(\delta^k\right)^2}} \cdot D \tag{14}$$

Finally, all agents must have an incentive to pick the prescribed ϵ^k (β^k). This requires that the indicated β^k minimizes

$$\left[[1 - p(\hat{q}_{i,n} = \bar{q}_n|\beta)] \cdot p^k - p[i\text{'s reports change } \bar{q}_n|\beta] \cdot Z + c(\beta)\right] \cdot N^k \tag{15}$$

Where the first term denotes the probability of being punished by getting the wrong signal, the second term denotes the probability of changing the market assessment of an object by one's own signal times the benefit of doing so ($Z \leq D$) and the third term denotes the cost of the chosen signal strength. By the proof of Theorem 1 that the first term in (15) is bounded above by

$$1 - p(\hat{q}_{i,n} = \bar{q}_n|\beta) \leq 1 - \beta_{\ell,\ell}\left(1 - 2e^{-2I^k\delta^2}\right) \leq 1 - \beta\left(1 - 2e^{-2I^k\delta^2}\right) \tag{16}$$

and the second term is bounded by $D\left(1 - 2e^{-2I^k\delta^2}\right)$.

I plug equation (16) in to equation (15) to continue the analysis. It will later become clear that using this upper bound is sufficient to show the result, i.e. the exponential terms disappear quickly in the limit. Therefore, the modified objective funciton is

$$\mathcal{V} = \left[\left[1 - \beta\left(1 - 2e^{-2I^k\delta^2}\right)\right] \cdot p^k + c(\beta)\right] \cdot N^k \tag{17}$$

Taking FOC of (17) and noting that that it must equal zero evaluated at $\beta = \beta^k$ it must be the case that

$$\frac{\partial \mathcal{V}}{\partial \beta}|_{\beta=\beta^k} = \left(1 - 2e^{-2I^k\delta^2}\right) \cdot p^k - c^{'}(\beta^k) = 0 \tag{18}$$

Rearranging,

$$p^k = \frac{c^{'}(\beta^k)}{1 - 2e^{-2I^k\delta^2}} \tag{19}$$

Taking a Taylor expansion of (19) yields

$$p^k = \frac{a_{\bar{d}-1} \cdot \left(\beta^k\right)^{\bar{d}-1} + h_{\bar{d}-1}(\beta^k) \cdot \beta^k}{1 - 2e^{-2I^k\delta^2}} \tag{20}$$

where $\bar{d}$ is the first non-zero derivative of $c(\cdot)$ at 0, and h is a residual term.

Therefore, a proposed plan $\left\{\epsilon^k, p^k\right\}_{k=1,2,\ldots}$ must asymptotically satisfy (11), (12), (14) and (20).

⇒**"if"** (11) is implied by

$$\epsilon^k > \left(\frac{\log N^k}{2I^k}\right)^{\frac{1}{2}-\gamma_1} \tag{21}$$

for any $0 < \gamma_1 < \frac{1}{2}$. Let $\epsilon^k = \left(\frac{1}{I^k}\right)^{\frac{1}{2}-\gamma_1}$ for some positive γ_1 sufficiently close to 0. From Taylor's theorem it is clear that in any plan $\left\{\epsilon^k, p^k\right\}_{k=1,2,\ldots}$ achieving the desired properties

$$p^k \to a_{\bar{d}-1} \cdot \left(\beta^k\right)^{\bar{d}-1} \tag{22}$$

By Lemma (3) we know that for large k

$$p^k \to 4^{\bar{d}-1} a_{\bar{d}-1} \left(\frac{1}{I^k}\right)^{(\frac{1}{2}-\gamma_1)(\bar{d}-1)} \tag{23}$$

Also note that this implies that (12) is satisfied when $(\frac{1}{2}-\gamma_1)(\bar{d}-1) \geq \alpha+1$, (where a similar Taylor expansion argument implies that $c(\epsilon^k) < p^k$ for sufficiently large k). Rearranging yields $d \leq 2\alpha+3$. All that remains is to show that (14) is satisfied, i.e. for sufficiently large k

$$4^{\bar{d}-1} a_{\bar{d}-1} \left(\frac{1}{I^k}\right)^{(\frac{1}{2}-\gamma_1)(\bar{d}-1)} \geq \frac{2e^{-2I^k\left(\left(\frac{1}{I^k}\right)^{\frac{1}{2}-\gamma_1}-\frac{1}{I^k}\right)^2}}{2\left(\frac{1}{I^k}\right)^{\frac{1}{2}-\gamma_1} - \left(2\left(\frac{1}{I^k}\right)^{\frac{1}{2}-\gamma_1}+1\right)2e^{-2I^k\left(\left(\frac{1}{I^k}\right)^{\frac{1}{2}-\gamma_1}-\frac{1}{I^k}\right)^2}} \cdot D \tag{24}$$

But note that $2\left(\frac{1}{2}-\gamma_1\right) < 1$ so the right hand side of (24) converges to zero exponentially in I^k while the left hand side converges at a much slower rate. Therefore, (14) is satisfied for sufficiently large k.

$\Leftarrow$**"only if"**

Note that (11) implies that $\epsilon^k > \left(\frac{2\log N^k}{I^k}\right)^{\frac{1}{2}}$ for sufficiently large k. But even for $\epsilon^k = \left(\frac{2\log N^k}{I^k}\right)^{\frac{1}{2}}$ (23) implies that $p^k \geq a_{\bar{d}-1}\left(\frac{2\log N^k}{I^k}\right)^2$ since $d \leq 2\alpha+3$. But then $\left(I^k\right)^{\alpha+1}\left(p^k\right) \to \infty$ violating (12).

Note that reworking the proof using equation (13) instead of (12) yields the complementary per capita cost result.

■

Proof of Point 1: The market designer need not consider the incentives of agents to lie, and so the designer must pick a sequence $\left\{\epsilon^k, p^k\right\}_{k=1,2,\ldots}$ to satisfy (11) and (12). Therefore, it is clear that setting $p^k = 0$ relaxes the constraints. Using a similar Taylor expansion argument as in the *Proof of 3.* for sufficiently large k

$$c(\epsilon^k) \approx a_{\bar{d}}(\epsilon^k)^{\bar{d}} \tag{25}$$

Therefore, in order to satisfy (12) and (21) it must be the case that $\frac{1}{2}\cdot\bar{d} > \alpha+1$ which is satisfied for any $\bar{d} \geq 2\alpha+2$. Similarly to the previous part, it is easy to see that if $\bar{d} < 2\alpha+2$ it is impossible to satisfy both (11) and (12), establishing the reverse direction of the claim.

Note that reworking the proof using (13) instead of (12) yields the complementary per capita cost result.

■

Proof of Proposition 2:

Recalling that agent n's maximum possible utility is D, she must be charged at least D or else she could (in some mechanisms, perhaps) lie and obtain a payoff of D. However, she must also account for the possibility that assesment $\bar{q}_{(n)}$ is incorrect and since there is always some small chance of this, she may lie.[16] By increasing the punishment by $2De^{-2I^k\delta^2}$ agent n no longer wishes to lie in any situation, as the equilibrium probability that $\bar{q}_{(n)} \neq q_{(n)}$ is bounded above by $2e^{-2I^k\delta^2}$ and the highest benefit to agent n is D. Therefore, in equilbrium, every object is correctly identified with probability 1. Since φ is strategy-proof, agents will report utilities and signals truthfully in mechanism $\bar{\varphi}^{CB}$. The claim regarding total expected punishment follows from Theorem 1 and the fact that in the limit no objects are incorrectly identified in equilibrium.

■

Proof of Proposition 3:

Agents still have incentives to report signals truthfully. By bounded-responsive preferences, each agent i need only report $v(x_{i.n}, n_q)$ for all $n \in N^k$ and $q_\ell \in Q$. In a strategy-proof ordinal mechanism φ, i need only report her ordinal rankings of $x_{i.n}$. When $q_{(N)}$ is not known, by continuity of v in its first argument agent i has no desire to misreport her relative preferences for n and n' if $|x_{i,n} - x_{i,n'}|$ is sufficiently large. To define "sufficiently large," begin by denoting $\gamma(\delta, I^k) = 2e^{-2I^k\delta^2}$, where γ represents the probability that $\bar{q}_{(n)} \neq q_{(n)}$. Then the expected utility of an agent i from misreporting preferences is bounded above by $D \cdot \gamma(\delta, I^k)$. On the other hand, by continuity of v the utility difference between $x_{i,n}$ and $x_{i,n'}$ is $\lambda \cdot |x_{i,n} - x_{i,n'}|$ for some $\lambda > 0$. Therefore, i has no incentive to misreport her preferences if

$$\forall n, n' \, \lambda \cdot |x_{i,n} - x_{i,n'}| < D \cdot \gamma(\delta, I^k).$$

Let where x^*be the value of x for which dF attains its maximum.Then the probability that i has an incentive to misreport is

$$\begin{aligned}
p\left(\min_{n,n'}|x_{i,n} - x_{i,n'}| < \frac{D}{\lambda} \cdot \gamma(\delta, I^k)\right) &\leq \\
\binom{N^k}{2} p\left(\min_{n,n'}|x_{i,n} - x_{i,n'}| < \frac{D}{\lambda} \cdot \gamma(\delta, I^k)\right) &\leq \\
\frac{\left(N^k\right)^2}{2} p\left(\min_{n,n'}|x_{i,n} - x_{i,n'}| < \frac{D}{\lambda} \cdot \gamma(\delta, I^k)\right) &\leq \\
\frac{\left(N^k\right)^2}{2} p\left(|x^* - x_{i,n}| < \frac{D}{\lambda} \cdot \gamma(\delta, I^k)\right) &\leq \\
\left(N^k\right)^2 \cdot g \cdot \frac{D}{\lambda} \cdot \gamma(\delta, I^k)
\end{aligned}$$

[16] Consider, for example, a situation in which n's utility is 0 in the allocation if she reports $\bar{s}_n = q_{(n)}$ and D if she reports $\bar{s}_n = q_\ell$ for some q_ℓ. Then since there is some chace that $\bar{q}_{(n)} = q_\ell$ she prefers to report $\bar{s}_n = q_\ell$.

The last inequality comes from the fact that dF is bounded above by g by assumption, and the probability of $x_{i,n}$ being within $\frac{D}{\lambda} \cdot \gamma(\delta, I^k)$ of x^* is bounded above by $2 \cdot g \cdot \frac{D}{\lambda} \cdot \gamma(\delta, I^k)$. Then the probability that any agent has an incentive to misreport preferences is bounded above by

$$I^k \left(\left(N^k\right)^2 \cdot g \cdot \frac{D}{\lambda} \cdot \gamma(\delta, I^k) \right).$$

The premise that $\left(N^k\right)^2 \in o\left(\frac{e^{2I^k\delta^2}}{I^k}\right)$ completes the truthtelling claim. Noting that the assumed growth rate of N^k is slower here than in Theorem 1 ensures that when $\left(N^k\right)^2 \in o\left(\frac{e^{2I^k\delta^2}}{I^k}\right)$ total waste converges to 0.

■

Proof of Proposition 4:

The particular scheme I consider is one in which the designer mixes equally between $S_i = 1$ for all i, $S_i = 0$ for all i and $S_i = 1$ independently with probability ρ for each i. For simplicity, I denote these S^1, S^0, and S^ρ respectively. Let $\epsilon^k = \epsilon\left(\frac{1}{I^k}\right)^t$ for some $t < 1$.[17] As $k \to \infty$ clearly $\epsilon^k \to 0$. This implies than the probability of ϵ^k-incorrectly identifying object n's characteristic under S^1, S^0, and S^ρ in any coordinated lie goes to zero generically. This follows from the fact that as $k \to \infty$ the measure of space in which objects are ϵ^k-identified goes to 0. The following remark states this formally.

Remark **1** *: For any $\eta > 0$, $\epsilon^k = \epsilon\left(\frac{1}{I^k}\right)^t$ and any coalition of agents $\hat{I}^k$ the probability of ϵ^k-incorrectly identifying an agent in S^1, S^0, and S^ρ by a coordinated misreport of signals by agents in $\hat{I}^k$ goes to zero for almost every $1 > \rho > 0$.*

Given this, the probability that an agent i places on S^1 (S^0) upon receiving $S_i = 1$ ($S_i = 0$) is $\frac{1}{1+\rho}$ $\left(\frac{1}{2-\rho}\right)$ by Bayes' rule. Therefore, choosing a punishment $p'(\rho) = (D+\iota) \cdot \max\left\{1 - \frac{1}{1+\rho}, 1 - \frac{1}{2-\rho}\right\} = (D+\iota) \cdot \max\left\{\frac{\rho}{1+\rho}, \frac{1-\rho}{1+\rho}\right\}$ for some $\iota > 0$ to be applied to all agents in the case that an object is unidentified implies that no coalition of agents wishes to misreport signals in mechanism $\bar{\varphi}$ for sufficiently large k. Note that for any $\iota > 0$, $p^{'}(\rho)$ is minimized by setting $\rho = \frac{1}{2}$, that is, when the designer makes different suggestions to different agents, she does so with equal probability.

The claim regarding total expected punishment follows from Theorem 1 and the fact that in the limit no objects are unidentified in equilibrium.

■

[17] I take t strictly less than 1 so that δ is well-defined in the definition of mechanism $\bar{\varphi}$.