Eliciting Information from a Large Population

Kohei Kawamura[*][†]
School of Economics, University of Edinburgh

This Version: 7 July 2011

Abstract

An uninformed welfare maximizing decision maker communicates with a random sample of agents from a large population who have heterogeneous preferences. Every agent's preference is private information to himself. The population distribution of preferences is unknown and messages from the sampled agents are cheap talk. This paper offers a tractable model of communication where the decision maker estimates the distribution of preferences through the messages, and chooses the policy to maximize welfare of the entire population. Information does not aggregate efficiently even if the sample size becomes arbitrarily large, since the sampled agents have incentive to "exaggerate" their preferences especially as the sample size becomes larger and each sampled agent has weaker influence on the decision. The quality of communication with each sampled agent may improve as the sample size becomes smaller, and thus we identify the trade-off between the quality and quantity of communication. We show that, given the same expected prior distribution of population preferences, the decision maker may prefer to sample a *smaller* number of agents when the prior is *weaker*.

Keywords: Sampling, Information Aggregation, Cheap Talk, Dirichlet Distribution, Aggregate Uncertainty

JEL Classifications: D71, D72, D82, H41

---

[*] I am grateful to Timothy Feddersen, Leonardo Felli, Gary Koop, David Myatt, Hikaru Ogawa, József Sákovics, Larry Samuelson, and seminar participants at Barcelona (UB), Edinburgh, Nagoya, Waseda, and WZB Berlin for helpful comments and discussions. All errors are my own.

[†] kohei.kawamura@ed.ac.uk

1 Introduction

Good public policy requires a considerable amount of information about the preferences of affected individuals. In many instances, authorities collect such information through consultation with representatives, polls of randomly selected individuals, or referendums. Private firms may also be interested in communicating with the public to find their "tastes" for marketing purposes. The process of information aggregation in those situations can potentially be complex, because the preference distribution may have to be estimated from data that is limited both in quality and quantity.

This paper explores the nature of communication for obtaining information about a large population. We introduce uncertainty about the *distribution* of preferences into a simple model that consists of an uninformed welfare maximizing decision maker and a continuum of agents with heterogeneous preferences. Specifically, we examine how the nature of cheap talk communication between the decision maker and randomly sampled agents changes according to the sample size and the quality of the prior belief about the preference distribution.

One of our main findings is the trade-off between quality of communication and sample size. Needless to say, if every sampled agent fully reveals their true preference, the decision maker is better off with a larger sample size as it renders the estimation of the distribution of preferences more accurate. However, as the sample size increases, each sampled agent has less influence on the decision maker's estimation of the preference distribution and consequently her decision. This leads to incentive to "exaggerate" their preferences, in the sense that if their type is high (low) they report that their type is even higher (lower) than it actually is, which implies that the quality of information transmission between the decision maker and each sampled agent diminishes as the sample size becomes larger. As a result, the decision maker cannot precisely infer the population distribution of preferences, even if the sample size is arbitrarily large. However, we also show that the quality of communication has a lower bound as there always exists an equilibrium where some information transmission occurs.

Another related finding, which is perhaps more interesting, is that the decision maker may be better off with sampling a *smaller* number of agents when the prior belief on the population preferences is *weaker*. This somewhat counterintuitive result comes from the fact that in communication with sampled agents, each agent plays not only against the other sampled agents but also against the decision maker's belief updating. If the prior belief is weak (i.e. if there is less ex ante information about the population distribution), the decision maker's estimation of the population distribution is influenced heavily by messages from the sampled agents. Therefore, each of them may have significant influence on the decision maker's belief and hence decision as long as the sample size is small, and thus they may reveal more information as they have less incentive to "exaggerate". This implies that a larger sample size does not necessarily lead to better estimation, and the optimal sample size may be bounded.

On the other hand, when the prior belief is strong and hence the decision maker is more confident about the preference distribution, a sampled agent has little influence on the decision maker's estimation (= Bayesian updating) of the population distribution even if the sample size is small, because a strong prior means that the prior and posterior distributions are likely to be similar. Consequently, each sampled agent has stronger incentive to "exaggerate" just as in the case of larger sample size, and the available quality of communication may hit the lower bound even if the sample size is, for instance, 1. The lower bound is binary communication ("yes or no", "for or against") so that sampled agents cannot possibly exaggerate. If the best available communication with each sampled agent is binary, then the decision maker is better off as the sample size becomes larger as she can better estimate the population distribution. This is in sharp contrast to the case where the prior is weak and the optimal sample size is bounded.

A recent paper by Morgan and Stocken (2008) studies information aggregation with cheap talk communication.[1] Similarly to our model, they consider a model with a decision maker and a continuum of agents with heterogeneous preferences. There are some important differences. First, Morgan and Stocken (2008) assume that the state and message spaces are binary, so that they do not analyze how the number of messages in equilibrium changes due to the agents' strategic incentive. Second, since they assume that the distribution of biases of the agents is known and aggregate uncertainty is only about the signals about a binary underlying state they receive, information typically aggregates as the size of the polls becomes arbitrarily large. This implies that if polling is costless, the decision maker always prefers to poll an arbitrarily large number of agents. In our model, on the other hand, information does not aggregate due to the complexity of the underlying state (, which is the preference distribution itself). A striking result that follows from the complexity is that even if polling is completely costless, the decision maker may prefer to sample a small number of agents because the quality of communication becomes higher. Therefore we are able to address the natural question of the optimal sample size, and moreover illustrate its relation to the quality of the prior distribution, which Morgan and Stocken (2008) do not examine.

Kawamura (2011) develops a similar setting to the one in the present paper, and studies how the finite number of interested parties (agents) affects information transmission from the agents. He also demonstrates the robustness of binary communication to exaggeration, but assumes that the decision maker communicates with all the agents involved, and hence does not address the issue of random sampling from a population. Most importantly the

---

[1] See e.g., Feddersen and Pesendorfer (1997, 1998) and Goeree and Großer (2007) for information aggregation in the context of voting.

trade-off between quality and quantity, which is the main focus of our analysis in the present paper, is absent in Kawamura (2011) since the average welfare is monotonically decreasing in the number of agents.

Other papers that study communication with several (mostly two) informed parties include Gilligan and Krehbiel (1987, 1989), Austen-Smith (1993), Krishna and Morgan (2001), Wolinsky (2002), Battaglini (2002, 2004), Ottaviani and Sørensen (2001) and Baliga, Corchon, and Sjöström (1997). They assume that the number of agents who communicate with the decision maker is given.

The intuition behind the quality-quantity trade-off we study in the present paper is somewhat related to that of the literature on committee design (Mukhopadhaya, 2003; Persico, 2004; Gerardi and Yariv, 2008; Gershkov and Szentes, 2009). In that literature, the optimal committee size is typically finite because committee members engage in costly information acquisition, which is subject to the free-rider problem even when the members have identical preferences. In our model, every agent (whether sampled or not) is endowed with private information which also represents their individual preferences.

Our technical contribution would be to introduce into a strategic setting the Dirichlet distribution to study random sampling under potentially complex population uncertainty.[2] The Dirichlet allows us to compute posteriors for a rich message structure (including partial pooling of types) and we can easily parametrize the strength of the prior belief, which is of interest in many practical situations.

The rest of the present paper proceeds as follows. The next section presents the model, and Section 3 examines the relationship among informative equilibria, sample size, and the quality of the prior. Section 4 considers the trade-off between quality and quantity of communication and the optimal sample size. Section 5 concludes.

2 Model

Our model consists of a single decision maker and a continuum of agents. Every agent is labelled by a real number $a \in [0,1]$ and has type $\theta_a \in \Theta \subset \mathbb{R}$. The number of types is $H \geq 3$, where the types are ordered such that $\theta^h < \theta^{h+1}$, for $h = 1, 2, ..., H$. The location for each type θ^h is fixed and common knowledge. We assume that the agents have a quadratic payoff function $-(y-\theta_a)^2$, where $y \in \mathbb{R}$ is the decision maker's policy. Clearly an agent's payoff is higher as y becomes closer to his type θ_a, and the "ideal" policy for

---

[2] See DeGroot (1970) pp.49-51, 174-175. We also use formulas from Dickey, Jiang and Kadane (1987). The Dirichlet distribution has been used in the economic literature for (non-strategic) search from an unknown distribution (e.g. Rothschild, 1974; Talmain, 1992).

agent a is $y = \theta_a$. The decision maker's objective is to maximize utilitarian "social welfare"

$$-\sum_{i=1}^{H} q^i (y - \theta^i)^2, \tag{1}$$

where $q^i \geq 0$ denotes the proportion of agents whose type is θ^i, such that $\sum_{i=1}^{H} q^i = 1$. Since y affects all the agents with different types, the decision maker is unable to implement the ideal policy for every individual (or type). Instead she chooses the policy that maximizes the total "welfare" of the agents' (1).

The decision maker does not observe the frequency vector $q \equiv (q^1, ..., q^H)$ or any agents' types directly. Each agent a learns only his own type θ_a. Therefore, the decision maker communicates with randomly sampled n agents to estimate q. In particular, the selected agents independently send cheap talk messages to the decision maker. In other words, communication is assumed to be completely costless and payoff irrelevant, and does not depend on n. We allow n to be any finite number, which can be arbitrarily large.[3]

We assume that the frequency vector q follows the Dirichlet distribution with parameters $\alpha \equiv (\alpha^1, \alpha^2 ..., \alpha^H) = (\alpha^0 p^1, \alpha^0 p^2 ..., \alpha^0 p^H)$ such that $\alpha^0 > 0$, $p^i > 0$ for all $i = 1, ..., H$, and $\sum_{i=1}^{H} p^i = 1$. It follows that $\sum_{i=1}^{H} \alpha^i = \sum_{i=1}^{H} \alpha^0 p^i = \alpha^0$. The density function with respect to the frequency vector q is given by

$$f(q; \alpha) = \frac{\Gamma(\alpha^0)}{\prod_{i=1}^{H} \Gamma(\alpha^i)} \prod_{i=1}^{H} \left(x^i\right)^{\alpha^i - 1},$$

where Γ denotes the gamma function. A convenient feature of the Dirichlet is that the marginal distribution of q^i is the beta distribution $B(\alpha^i, \alpha^0 - \alpha^i)$. This implies that the prior expectation of the frequency of the ith type is

$$E[q^i] = \frac{\alpha^i}{\alpha^0} = \frac{\alpha^0 p^i}{\alpha^0} = p^i. \tag{2}$$

Therefore we can see $p \equiv (p^1, p^2, ..., p^H)$ as the expected prior population distribution of preferences. The decision maker and all agents share the same prior, before the agents learn their types. In our common prior Bayesian framework, each agent updates his belief on the population distribution through his own type, while the decision maker updates her belief on the population distribution through n messages she receives from the sampled agents. Let the prior expected mean of the agents' types be

$$\mu \equiv \sum_{i=1}^{H} p^i \theta^i.$$

---

[3] We rule out the case where n is infinitely large, so that each sampled agent's message has some (however small) influence on the decision maker's belief and policy.

For expositional convenience we assume $\mu \neq \theta^i$ for all $i = 1, ..., H$. In other words, we rule out the non-generic case where the prior mean coincides exactly with one of the types.[4] Suppose that every sampled agent reveals their type truthfully. Let $x = (x^1, ..., x^H)$ be the count vector where x^i denote the number of agents whose type is θ^i, out of n sampled agents. Clearly we have $\sum_{i=1}^{H} x^i = n$. From the decision maker's viewpoint the posterior distribution of q^i is the beta distribution $B(\alpha^i + x^i, \alpha^0 - \alpha^i + n - x^i)$ and the expected frequency q^i conditional on x^i is given by

$$E[q^i \mid x^i] \; = \frac{\alpha^0 p^i + x^i}{\alpha^0 + n}. \tag{3}$$

This reflects the convenient property of the Dirichlet distribution that the posterior of q^i is affected only by x^i and the sample size n, and not by the count of any other agent x^{-i}.

From each agent's viewpoint, after he learns his type $\theta_a = \theta^i$, the posterior distribution of the probability mass of his type, q^i, is $B(\alpha^i + 1, \alpha^0 - \alpha^i)$. That of any other type, denoted by q^{-i}, is given by $B(\alpha^{-i}, \alpha^0 - \alpha^{-i} + 1)$. Hence we obtain

$$\begin{cases} E[q^i \mid \theta_a = \theta^i] \; = \frac{\alpha^0 p^i + 1}{\alpha^0 + 1} \\ E[q^{-i} \mid \theta_a = \theta^i] \; = \frac{\alpha^0 p^{-i}}{\alpha^0 + 1} \end{cases} \tag{4}$$

That is, (4) describes the expected posterior distribution of the population preferences from the viewpoint of an agent whose type is θ^i. Note that each agent updates his belief according to the sample size of 1, which is his own type.

The Dirichlet distribution is used widely in problems where the underlying distribution is unknown. It provides a tractable way to model a "distribution of distributions". By construction, the expected prior distribution p can be completely arbitrary. Furthermore, we can parametrize the strength of the prior by α^0. As we can see from (3) and (4), the higher α^0 is, the less "sensitive" the posterior is with respect to the types of the sampled agents because

$$\frac{\Delta E[q^i \mid x^i]}{\Delta x^i}$$

is decreasing in α^0. Also, α^0 can also be seen as the level of ex ante aggregate uncertainty, conditional on prior common knowledge about the population distribution. When α^0 is high, the realized population distribution is likely to be similar to the prior. For example, if $\alpha^0 \to \infty$ then $E[q^i \mid x^i] \to E[q^i] = p^i$ for any x^i. In this case, the prior is identical to the posterior (and hence the realized population distribution) with probability 1, which corresponds to a completely known population distribution (i.e. no aggregate uncertainty). Consequently the decision maker can choose the (near) first-best policy even without any communication. In contrast, when α^0 is a finite number, the realized distribution may

---

[4] We will discuss the implication of this assumption later.

well be different from the prior and there is uncertainty in the population distribution of preferences.

From an agent's perspective, the other agents' types are correlated with his own since (4) implies $E[q^i \mid \theta_a = \theta^i] > E[q^i]$ for finite α^0. The level of correlation is decreasing in α^0, as we have

$$\frac{dE[q^i \mid \theta_a = \theta^i]}{d\alpha^0} < 0.$$

In other words, the lower α^0 is, the more likely the others are of his type. In particular, if $\alpha^0 \to 0$ we have $E[q^i \mid \theta_a = \theta^i] \to 1$, which means the other agents' types are perfectly correlated with his (i.e. all the others share the same type as his own). Thus aggregate uncertainty implies correlation of types (and vice versa) in the present framework.[5][6]

The timing of the game is as follows:

1. All agents and the decision maker are endowed with a common prior on the preference distribution;
2. Agents privately learn their types;
3. The decision maker randomly samples n agents who report costless, non-verifiable messages;
4. The decision maker estimates the population distribution from the messages and chooses y;
5. Payoffs are realized.

In what follows we introduce the possibility that the agents may not fully reveal their types. In particular, we will see that the type space may be partitioned. Note that while the decision maker has n pieces of information (messages), each agent has only one (his own type).

3 Equilibrium

Throughout this paper we focus on common partitional strategies for the agents, where each non-overlapping group consists of one or more type indices in row and any sampled

---

[5] The negative association between the strength of the prior and correlation follows directly from the Dirichlet assumption. This need not be the case in general, but is likely to hold if the observation of a type increases the expected frequency of this type, and reduces that of the others.

[6] The covariance of any two different types is given by $\frac{-\alpha^i\alpha^{-i}}{(\alpha^0)^2(\alpha^0+1)}$, so that any type is negatively correlated with the other types. In other words, from an agent's viewpoint the correlation is with respect to his own type only, and not to any others. This suggests that each type in our discrete type space could be better interpreted as a simplified representation of (possibly continuous) types that are close to each other and positively correlated.

agents in the same type group induce (from their viewpoint) the same distribution of policy by the decision maker.[7] Without loss of generality we assume that all sampled agents in the same group send the identical message.[8]

Let the number of the groups on the type space be $K \leq H$. Let G^k be the set of type indices in the kth group from the right hand side of the type space. For example, if $K = H$ each type reports a distinct message to the decision maker, and $G^k = \{k\}$. On the other hand, if $K = 1$, then G^1 contains all types: $G^1 = \{1, 2, .., H\}$. Let $z \equiv (z^1, ..., z^k, ..., z^K)$ be the count vector of messages from the sampled agents in each group. Naturally we have $\sum_{k=1}^{K} z^k = n$. The first order condition with respect to (1) gives the decision maker's best response conditional on the messages:

$$\bar{y}(z) = \sum_{i=1}^{H} E[q^i \mid z] \times \theta^i, \tag{5}$$

where $E[q^i \mid z]$ is the posterior expected frequency of type i.

Let $G(i)$ denote the set of type indices (group) that has i as an element. By definition, if $i+1 \in G(i)$ then $G(i) = G(i+1)$. If type i is in the kth group, then $G(i) = G^k$. Suppose that all types in $G(i)$ send the same message to the decision maker and hence she cannot tell exactly how many of the sampled agents are of type θ^i. For the type of agent θ^i we have

$$E[q^i \mid z] = \underbrace{\frac{\alpha^0 \sum_{l \in G(i)} p^l + z(i)}{\alpha^0 + n}}_{\text{"pooled" expected frequency}} \underbrace{\frac{p^i}{\sum_{l \in G(i)} p^l}}_{\text{weight within } G(i)}, \tag{6}$$

where $z(i)$ denotes the number of sampled agents in $G(i)$, i.e. $z(i) \equiv \sum_{l \in G(i)} x^l$.[9] Note that the decision maker does not observe x^i directly for $G(i)$ that contains two or more types.

From each sampled agent's viewpoint, his message induces a distribution of the decision maker's policy, which is influenced also by the other sampled agents' messages. Note from (5) and (6) that the count vector z for messages influences the decision maker's policy linearly. This implies that from each sampled agent's viewpoint his message influences the expectation but not the variance of the policy. Therefore, as we assume quadratic payoffs,

---

[7] In principle there might be other types of non-partitional equilibrium. We are unable to provide a full equilibrium characterization since the decision maker's posterior then does not have a closed-form solution. We will discuss this later in this paper.

[8] Sampled agents in the same group do not have to send the identical message, as long as they induce the same probability distribution of policy (or equivalently the same belief of the decision maker on their types). However, such an equilibrium is outcome equivalent to the one where all agents in the same group send an identical message, in the sense that the same combination of the sampled agents' types results in the same policy.

[9] See Dickey, Jiang and Kadane (1987), pp777-780.

we can focus on the expected policy induced by each message when we consider a sampled agent's strategy.

Every agent updates his own belief on the population distribution, according to his own type. Given that the agent's type is θ^i, the types of all the other agents are Dirichlet distributed with parameters $\alpha' = (\alpha^1, ..., \alpha^i + 1, ..., \alpha^H)$. This implies that the posterior distribution of his own type is $B(\alpha^i+1, \alpha^0-\alpha^i)$ while that of the other types is $B(\alpha^{-i}, \alpha^0 - \alpha^{-i}+1)$, where θ^{-i} denotes a type other than θ^i.

If the underlying preference distribution is Dirichlet distributed, the count vector x follows the multivariate Pólya distribution (also known as the Dirichlet compound multinomial distribution):

$$\Pr(x \mid \alpha) = \frac{n!}{\Pi_{i=1}^{H}(x^i!)} \frac{\Gamma(\alpha^0)}{\Gamma(\alpha^0+n)} \prod_{i=1}^{H} \frac{\Gamma(\alpha^i + x^i)}{\Gamma(\alpha^i)},$$

where Γ is the gamma function and x^i is the number of sampled agents whose type is θ^i.

Let us consider the sampled agents' choice of messages. We denote the message sent by any agents in $G(i)$ by $m(i)$; and the message sent by any agents in G^j by m^j. If a sampled agent's type is θ^i and he sends the message $m(i)$, then his expected payoff is given by

$$\begin{aligned} &u(\theta^i, m(i)) = \\ &\quad -\sum_{x^1=0}^{n} \sum_{x^2=0}^{n-x^1} \cdots \sum_{x^H=0}^{n-x^1-,...,-x^{H-1}} \Pr(x \mid \alpha') \left(\theta^i - \sum_{t=1}^{H} E[q^t \mid z^1, ..., z(i)+1, ..., z^K] \times \theta^t \right)^2, \end{aligned}$$

where $z^k = \sum_{l \in G^k} x^l$ for $k = 1, ..., K$. If he deviates and mimics a type in the jth group such that $G^j \neq G(i)$, then

$$\begin{aligned} &u(\theta^i, m^j) = \\ &-\sum_{x^1=0}^{n} \sum_{x^2=0}^{n-x^1} \cdots \sum_{x^H=0}^{n-x^1-,...,-x^{H-1}} \Pr(x \mid \alpha') \left(\theta^i - \sum_{t=1}^{H} E[q^t \mid z^1, ..., z^j+1, ..., z(i), .., z^K] \times \theta^t \right)^2. \end{aligned}$$

Note that the expected payoff function $u(\theta^i, \cdot)$ already incorporates (5), the decision maker's best response to the messages from the sampled agents given that all of them follow the partitional strategy. Hence a perfect Bayesian equilibrium is such that for any θ^i,

$$u(\theta^i, m(i)) \geq u(\theta^i, m^j) \quad \forall m^j = m^1, m^2, ..., m^K.$$

The expected payoff function $u(\theta^i, \cdot)$ is complex, but since the original payoff function is quadratic, for each agent's best response conditional on his type, we can focus on the decision maker's expected policy (from an agent's viewpoint) induced by his message. Specifically, it suffices to find which message induces the expected policy closest to his ideal policy θ_a.

How does communication between the decision maker and the sampled agents take place in equilibrium? The following proposition states that, when the prior belief about the preference distribution is strong enough, generically the only informative equilibrium communication is the one that can be played with binary messages (e.g. "yes or no"). In this equilibrium, the agents' types are partitioned into only two groups, and any sampled agent from a group induces the same belief as the other agents in the group. In other words, the decision maker can correctly infer to which of the two type groups a sampled agent belongs, but cannot precisely know the agent's type.

Proposition 1 For α^0 sufficiently large, the only informative equilibrium is binary for any n, whereby all types below μ belong to one group and above μ belong to the other.

Proof. See Appendix I. ■

As we have already suggested in the Introduction, this proposition has a simple intuition. When the prior is stronger, each sampled agent has a smaller influence on the decision maker's belief and hence her policy, regardless of the sample size. Also, his expected prior policy becomes closer to the prior expectation μ. As a result, a sampled agent has incentive to exaggerate his type relative to his posterior expected policy (a point close to μ) by the decision maker, in order to render the policy y closer to his ideal θ_a. In other words, agents whose types are above μ wish to overstate their types as much as they can, and agents below μ understate their types as much as they can. Binary communication is "robust" to this exaggeration incentives, because the sampled agents may not possibly exaggerate their types when they have the choice between two messages (above or below μ).

A larger sample size has a similar effect on communication to a stronger prior, in the sense that it weakens the decision maker's response to each sampled agent's message.

Proposition 2 For n sufficiently large, either i) no informative equilibrium in partitional strategy exists; or ii) the most informative equilibrium in partitional strategy is binary.

Proof. See Appendix I. ■

When a large sample size is combined with a moderate or strong prior, it leads to binary communication for essentially the same reason and intuition as in Proposition 1.[10]

---

[10] Kawamura (2011) has given a related proposition but with a finite number of agents all of whom send a message to the decision maker (hence there is no sampling), where the decision maker concerns only the types of those agents, not the underlying probability distribution of types itself. In the present framework the decision maker's Bayesian updating is much more complex because she has to estimate the entire population distribution regardless of the sample size n. In other words, the decision maker has to assign a (strictly positive) posterior probability mass to *all* possible types $\theta^1, \theta^2, ..., \theta^H$ even when the sample size is very small or when no sampled agent turns out to be of certain types.

However, if the prior is extremely weak, even binary communication equilibrium may not exist for a very large sample size since types close to the boundary types in the binary partition has incentive to deviate.

Note that when α^0 is close to 0 a large proportion of the population are likely to be concentrated on one type due to high correlation. Consider the highest type in the lower group of a binary partition. From the viewpoint of a sampled agent who finds himself having this type, the expected action (given that the other sampled agents follow the binary partitional strategy) will be *lower* than his ideal, since the other sampled agents are very likely to share the type and induce the decision maker's belief that their expected type is the expected type of the lower group, not the highest type in the group. Thus, the sampled agent may mimic a type in the *higher* group to render the expected policy higher, which upsets the binary equilibrium.

This does not imply that there is no informative equilibrium. In fact, even for small α^0 and large n there could be a mixed strategy equilibrium where sampled agents randomize their messages. Unfortunately it is impossible to characterize a mixed strategy equilibrium since the posterior is no longer the Dirichlet and hence does not have a closed form.[11] However, in our framework very small α^0 has a somewhat unrealistic feature that, due to high correlation, a vast majority of the agents is likely to be concentrated on one type, which the decision maker does not know. In practice, such situations may be of less interest because there is little intrinsic conflict of interest between the decision maker and the agents. In what follows we focus on α^0 such that a binary equilibrium exists, which implies we expect see some dispersion of realized preferences.

It is easy to see that, given binary communication, the decision maker can never estimate the preference distribution precisely. Figure 1 shows an example of two distributions that the decision maker is unable to distinguish even if the sample size is arbitrarily large: she can almost precisely estimate the proportions of the agents are below and above μ, but she can never accurately infer how the types are distributed above μ. This makes it impossible for the decision maker to implement the first best policy for any sample size.

Note that as in most cheap talk models, there always exists the "babbling" equilibrium where the sampled agents uninformative messages and the decision maker's policy is the same as the one based only on the prior. However, it is clear that the decision maker is strictly better off in an informative equilibrium as she can use her information to maximize her conditional expected payoff (therefore for any realization of types she cannot be worse off). This also implies that all the agents are ex ante better off in an informative equilibrium.

Propositions 1 and 2 assumes that there does not exist a type that coincides with the

---

[11] Moreover, as there is no guarantee that messages shift the policy linearly, we cannot focus on the expectation of y and have to take into account the variance, which also makes our analysis intractable.

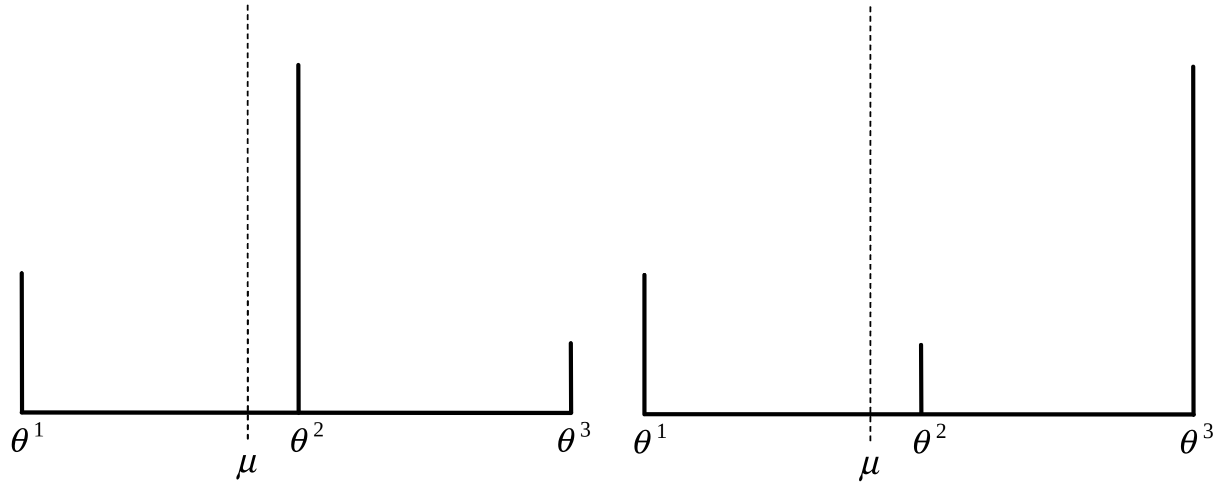

Figure 1: Two observationally equivalent preference distributions for binary data

prior expected mean μ. If such a type exists, this type does not have incentive to exaggerate as his posterior expectation on the other agents' types also coincides with his type and the most informative equilibrium features three groups, not two. Clearly, when $H \geq 4$ even if we have ternary communication where with all agents whose type coincides with μ sending a distinct message, there does not exist a fully revealing equilibrium for large n or α^0.

4 Sample Size and Quality-Quantity Trade-off

In the previous sections we have seen that the sample size may be negatively associated with the quality of each message. Meanwhile, it is clear that given the quality (informativeness) of communication between the decision maker and each agent, a larger sample size allows the decision maker to estimate the population distribution more accurately. This suggests an interesting trade-off between quality and quantity of messages from sampled agents.

When the underlying uncertainty on the preference distribution simple, even coarse communication may allow the decision maker to identify the population distribution precisely, as the sample size becomes arbitrarily large. This is the case, for example, when the population distribution is normal with an unknown mean, where the exact proportion of the agents below or above a threshold type gives sufficient information to precisely infer the entire distribution. The Dirichlet has much less structure on its posterior. Hence in order for the decision maker to estimate the population distribution exactly, every agent must reveal truthfully and also the sample size must be arbitrarily large. However, from Proposition 2 we know that this *cannot* be an equilibrium outcome.

In the following we study how the decision maker's ex ante expected equilibrium payoff ("social welfare") changes according to the number of sampled agents, by assuming that

$\alpha^0 = 7/5$

n	perfect	binary
0	-0.0898	
1	$*-0.0742$	-0.0788
2	$*-0.0678$	-0.0742
3	$*-0.0643$	-0.0717
4	$*-0.0621$	-0.0702
5	$*-0.0606$	-0.0691
6	N/A	-0.0683
$\to \infty$	N/A	$*-0.0633$

$\alpha^0 = 4$

n	perfect	binary
0	-0.0898	
1	N/A	-0.0873
2	N/A	-0.0856
3	N/A	$*-0.0844$
4	N/A	$*-0.0835$
5	N/A	$*-0.0828$
6	N/A	$*-0.0822$
$\to \infty$	N/A	$*-0.0771$

Table 1: Decision maker's expected payoff (= "social welfare") when $\theta^1 = 0, \theta^2 = 1/2, \theta^3 = 0, p^1 = 2/8, p^2 = 5/8, p^3 = 1/8$. An asterisk(*) denotes social welfare in neologism proof equilibrium.

the decision maker can commit to a sample size and the sampled agents know it. This is an important assumption in our model not least because conditional on a certain number of received messages from the sampled agents, the decision maker is always tempted to sample more to estimate the population distribution better. Consequently, knowing that the decision maker cannot commit to an announced sample size, the sampled agents would anticipate that the actual sample size is arbitrarily large, in which case only binary communication is available. On the other hand, if the decision maker has to choose her policy immediately after communication (and no time is left for a second round communication) then this time constraint itself may function as a commitment device.

The decision maker's expected payoff conditional on the sample size n and the best response (5) is computed by

$$u^{DM}(n) = -\sum_{x^1=0}^{n} \sum_{x^2=0}^{n-x^1} \cdots \sum_{x^H=0}^{n-x^1-,\dots,-x^{H-1}} \underbrace{\Pr(x \mid \alpha)}_{\text{dist. of type counts}} \underbrace{\left(\sum_{i=1}^{H} E[q^i \mid z] \left(\bar{y}(z) - \theta^i \right)^2 \right)}_{\text{exp. payoff conditional on message counts}}, \quad (7)$$

where $\Pr(x \mid \alpha)$ denotes the multivariate Pólya distribution, $z = (z^1, .., z^K)$ is the count vector of the messages in a partitional equilibrium and $z^k = \sum_{l \in G^k} x^l$ as described above.

Let us observe the quality-quantity trade-off through an example, the details of which can be found in Appendix II. Table 1 presents the decision maker's ex ante expected payoff (i.e. ex ante "social welfare") according to sample size n when $H = 3$ ($\theta^1 = 0$, $\theta^2 = 1/2$, $\theta^3 = 1$) and the prior expected mean $\mu = 7/16 < 1/2$. In this example the middle type θ^2, which is above the expected mean, has incentive to exaggerate and mimic the high type θ^3. There are multiple equilibria in this game, and there always exists a "babbling" equilibrium where all sampled agents send uninformative equilibrium. The payoff in the babbling equilibrium is the same as that of no communication at all $n = 0$. The most informative equilibrium in partitional strategy features either perfect communication where each type θ^i for $i = 1, 2, 3$ sends a distinct message, or binary where agents with θ^1 and θ^2 send the same message and those with θ^3 send a separate message.

We can see that, if $\alpha^0 = 7/5$ and the most informative equilibrium is chosen, the social welfare is non-monotonic in the sample size since the welfare under perfect communication for $n = 5$ (-0.0606) is higher than the welfare under binary communication for $n = 6$ (-0.0683).[12] Moreover, the social welfare for $n = 5$ is higher than for $n \to \infty$, which implies that the optimal sample size is bounded even if sampling itself is completely costless. This is because for $n \geq 5$ the incentive to exaggerate is so strong for the sampled agents that, there does not exist an equilibrium where every type reveals truthfully. When the prior is stronger (the expected payoffs are listed for $\alpha^0 = 4$), each sampled agent has weaker influence on the decision maker's posterior hence her policy as we have seen in (3). This leads to larger incentive to exaggerate even if the sample size is very small. Consequently the only informative partitional equilibrium is binary, regardless of the sample size. Therefore, the expected social welfare is monotonically increasing in the sample size n.

The two cases in Table 1 indicate that, given the same expected prior distribution, the decision maker may prefer to sample a *smaller* number of agents when the prior is *weaker*. Our discussion in this section can be summarized as follows.

Proposition 3 If α^0 is sufficiently large, the optimal sample size is unbounded. Otherwise, the relationship between social welfare and n may be non-monotonic and the optimal sample size may be bounded.

Our analysis so far has focused on sampled agents using partitional strategies, where every type in a group sends essentially the same message with probability 1. There may also be a mixed strategy equilibrium, which unfortunately is impossible to characterize since the posterior does not have a closed form. However, we can provide some arguments for the partitional equilibria we have focused on and for the claim that the optimal sample size can be bounded. First, for smaller sample size, if α^0 is not too high and there exists a fully revealing equilibrium, it must be ex ante more efficient than the others for given n, since more precise information is transmitted to the decision maker than binary or possibly a mixed strategy equilibrium. Furthermore, the fully revealing equilibrium is robust in the sense that it is "neologism proof" (Farrell, 1993), so that no sampled agent can be better off by sending credible off-the-equilibrium messages.[13]

Binary communication is not neologism proof for relatively small α^0 and n, as in the above example, because then a high type agent θ^3, who always wants to render the policy as

---

[12] We assume that all sampled agents play the same equilibrium strategy.

[13] See Appendix II for a detailed discussion on neologism proofness in the example of this section.

high as possible, may have a neologism (off-the-equilibrium message) to credibly distinguish himself from the middle type θ^2. The middle type will not use the neologism since, due to relatively small α^0 and n, it has a large influence on the decision maker and makes the expected policy too high for him, which in turn makes the neologism by the high type credible. On the other hand, as long as binary communication equilibrium exists for $n \to \infty$, as in Table (1), it is neologism proof for large n because a middle type agent who now has weak influence on the policy would also use the neologism by a high type agent, which means no credible neologism exists. In other words, for large n the binary equilibrium is not only simple but also robust.

5 Concluding Remarks

This paper has proposed a novel approach to analyze cheap talk communication to elicit information on a large population. Due to the presence of aggregate uncertainty and sampled agents' incentive to misreport their preferences, the welfare maximizing decision maker has to estimate the preference distribution. Since a large sample size may diminish the quality of communication with each sampled agent, the optimal sample size may be bounded, even if communication and information processing are completely costless. This is particularly the case when the prior belief on the population distribution of preferences is *weak*.

Throughout this paper we have assumed the decision maker can commit to a sample size. Perfect commitment to a sample size may seem contradictory to the assumption that the decision maker optimally responds to messages without committing to a mechanism. However, if the decision maker has to choose her policy immediately after communication (and no time is left for a second round communication) then this time constraint itself may function as a commitment device: the decision maker may credibly sample a fixed number of agents to ask their preferences.

For a decision maker who is not time constrained, a natural extension of our model is sequential sampling. In this case, the commitment problem seems severer because the decision maker will always be tempted to ask more agents, as long as communication is costless and there is no time constraint. We could introduce a cost of sampling, in which case the decision maker will determine when to stop sampling, depending on the information she has obtained. If the decision maker can set the cost of sampling, it may become a commitment device to sampling a small number of agents. We would then have to give up our parametric Bayesian approach as there is no known analytical solution for the optimal stopping problem when the type space is non-binary.

6 Appendix I

6.1 Proposition 1

Proof. Consider an arbitrary partition of the type space $\{\theta^1, \theta^2, ..., \theta^H\}$ into $J(\leq H)$ disjoint groups. Let $\theta^{(j,1)}$ be the lowest and $\theta^{(j,S_j)}$ be the highest type in the jth group that consists of S_j types. Let $\pi : (j,s) \mapsto \{1, 2, ..., H\}$ be a function from the identity of a group j and the order within the group s to the original type ordering. Clearly we have $\pi(1,1) = 1$ (i.e., $\theta^{(1,1)} = \theta^1$) and $\pi(J, S_J) = H$ (i.e., $\theta^{(J,S_J)} = \theta^H$). For a group with a single type, $S_j = 1$ and $\theta^{(j,S_j)} = \theta^{(j,1)}$. In the following we will show that for any partition such that $J \geq 3$ and $\theta^{(2,1)} < \mu$ cannot be an equilibrium if α^0 is large enough, since this boundary type $\theta^{(2,1)}$ has incentive to deviate and mimic a type in the first group. Likewise, for any partition such that $J \geq 3$ and $\theta^{(J-1,S_{J-1})} > \mu$, this type $\theta^{(J-1,S_{J-1})}$ deviates and mimics a type in the Jth group.

First, let us denote the expected type of an agent in the jth group by

$$\bar{\theta}(j) \equiv \sum_{g=\pi(j,1)}^{\pi(j,S_j)} \frac{\alpha^g}{\sum_{s=1}^{S_j} \alpha^{(j,s)}} \theta^g.$$

Note that $\bar{\theta}(j)$ depends only on the parameters of the prior and is independent from the ex post realization of agent types. Thus the variance of the decision maker's policy from the agent's viewpoint does not change according to his individual message, and thus we can focus on which message induces the closest expected policy to the agent's ideal policy.

Regardless of the partition of the type space, the ex ante expected type of the agents from the decision maker's viewpoint is μ. In other words,

$$\sum_{j=1}^{J} \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)}}{\alpha^0} \bar{\theta}(j) = \sum_{i=1}^{H} \frac{\alpha^i}{\alpha^0} \theta^i = \mu.$$

Suppose that a sampled agent has learnt his type, and let us consider how the other sampled agents affect the decision maker's belief (and policy). Note that the partition of types plays an important role because the decision maker's Bayesian updating is based on it. Let z^j be the number of sampled agents in group j. We have $\sum_{j=1}^{J} z^j = n$. If all sampled agents follow the partitional strategy, the decision maker's policy conditional on their messages is given by

$$y(z^1, z^2, ..., z^J) = \hat{\mu}(z^1, z^2, ..., z^J) = \sum_{j=1}^{J} \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)} + z^j}{\alpha^0 + n} \bar{\theta}(j).$$

Suppose that all sampled agents except θ_a follows the partitional strategy and $\theta_a = \theta^{(2,1)}$ (i.e. he is the the lowest type in the second group). If this sampled agent follows the partitional strategy, the expected policy of the decision maker is given by

$$\begin{aligned}
\bar{\mu}_2(2) &\equiv \sum_{j=1}^{J} \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)} + E[z^j \mid \theta_i = \theta^{(2,1)}]}{\alpha^0 + n} \bar{\theta}(j) + \frac{1}{\alpha^0 + n} \bar{\theta}(2) \\
&= \sum_{j=1, j \neq 2}^{J} \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)} + (n-1) \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)}}{\alpha^0+1}}{\alpha^0 + n} \bar{\theta}(j) + \frac{\sum_{s=1}^{S_2} \alpha^{(2,s)} + (n-1) \frac{\sum_{s=1}^{S_2} \alpha^{(2,s)}+1}{\alpha^0+1} + 1}{\alpha^0 + n} \bar{\theta}(2) \\
&= \frac{\alpha^0}{\alpha^0 + 1} \mu + \frac{1}{\alpha^0 + 1} \bar{\theta}(2). \qquad (8)
\end{aligned}$$

Since this is a convex combination of the prior expectation of agent types μ and the expected type of the group where the agent belongs to, if $\theta^{(2,1)} < \mu$ then $\theta^{(2,1)} < \bar{\mu}_2(2) < \mu$.

If the sampled agent with $\theta_a = \theta^{(2,1)}$ mimics an agent in the first group, then the expected action of the decision maker is

$$\begin{aligned}
\bar{\mu}_2(1) &= \frac{\sum_{s=1}^{S_1} \alpha^{(1,s)} + E[z^1 \mid \theta_i = \theta^{(2,1)}] + 1}{\alpha^0 + n} \bar{\theta}(1) + \sum_{j=2}^{J} \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)} + E[z^j \mid \theta_i = \theta^{(2,1)}]}{\alpha^0 + n} \bar{\theta}(j) \\
&= \sum_{j=1}^{J} \frac{\sum_{s=1}^{S_j} \alpha^{(1,s)}}{\alpha^0 + 1} \bar{\theta}(j) + \frac{1}{\alpha^0 + n} \bar{\theta}(1) + \frac{(n-1)}{(\alpha^0 + n)(\alpha^0 + 1)} \bar{\theta}(2) \\
&= \bar{\mu}_2(2) - \frac{1}{\alpha^0 + n} (\bar{\theta}(2) - \bar{\theta}(1)). \qquad (9)
\end{aligned}$$

We now observe that for large enough α^0

$$\theta^{(2,1)} < \bar{\mu}_2(1) < \bar{\mu}_2(2), \qquad (10)$$

which implies that the decision maker's policy is closer to his ideal when he mimics an agent in the first group whose expected type is lower. Note that from (5) and (6) the agent's message does not influence the variance of the decision maker's policy. Thus (10) implies that the expected payoff of an agent in the second group is higher if he mimics one in the first group.

Similarly, consider a sampled agent whose type is the largest in the $(J-1)$th group: $\theta_a = \theta^{(J-1,S_J-1)}$. If $\theta^{(J-1,S_J-1)} > \mu$ then $\theta^{(J-1,S_{J-1})} > \bar{\theta}(J-1)$. Hence $\theta^{(J-1,S_{J-1})} > \hat{\mu}_2$, which means that from the viewpoint of the agent $\theta^{(J-1,S_{J-1})} > \mu$, the decision maker's expectation on any other agent is lower than his own type. Therefore, for large enough α^0, the decision maker's policy is closer to his ideal policy when he mimics an agent in the Jth group:

$$\theta^{(J-1,S_{J-1})} > \bar{\mu}_{J-1}(J) > \mu_{J-1}(J-1). \qquad (11)$$

Hence for any arbitrary partition, if $\theta^i \neq \mu$ for all i and there exists a type of agent who does not belong to the first or the last group, he has incentive to deviate when α^0 is large enough.

The above argument rules out any partitional equilibrium with three or more groups. Consider a binary partition where there are only two groups and $\theta^{(1,S_1)} < \mu < \theta^{(2,1)}$. No agent deviates for large α^0 because mimicking a type in the other group renders the expected policy further away from the ideal:

$$\theta^1 < ... < \theta^{(1,S_1)} < \bar{\mu}_1(1) = \frac{\alpha^0}{\alpha^0+1}\mu + \frac{1}{\alpha^0+1}\bar{\theta}(1) \tag{12}$$

and

$$\bar{\mu}_1(2) = \frac{\alpha^0}{\alpha^0+1}\mu + \frac{1}{\alpha^0+1}\bar{\theta}(2) < \theta^{(2,1)} ... < \theta^H. \tag{13}$$

Hence we conclude that the only informative equilibrium is binary for large enough α^0. ■

6.2 Proposition 2

Proof. From (9), (10) and also (11) hold for large enough n, regardless of α^0. This rules out any partitional equilibrium with three or more groups.

On the other hand, a binary partition equilibrium may not exist either if n is large and α^0 is close to 0. To see this, consider any binary partition where the second group contains two or more types. Then for α^0 close to 0

$$\theta^{(2,1)} < \bar{\mu}_2(2) = \frac{\alpha^0}{\alpha^0+1}\mu + \frac{1}{\alpha^0+1}\bar{\theta}(2) < \bar{\theta}(2).$$

Then from (9) we have $\theta^{(2,1)} < \bar{\mu}_2(1) < \bar{\mu}_2(2)$ in the binary partition when n is large, which implies that a sampled agent with $\theta^{(2,1)}$ mimics an agent in the first group. The same argument holds for the first group if it has two or more types. Therefore, an equilibrium in partitional strategy does not exist if n is large enough and α^0 is close to 0. ■

7 Appendix II

In this appendix we provide details of the example in Section 4, where $\theta^1 = 0$, $\theta^2 = 1/2$, $\theta^3 = 1$ and the expected prior for each type is given by $p^1 = 2/8$, $p^2 = 5/8$, $p^3 = 1/8$. The prior mean $\mu = 7/16$ and hence from the viewpoint of the middle type θ^2, fully revealing communication biases the policy lower than the ideal.

First let us consider the condition under which the binary communication equilibrium exists for any n. Note that, given θ^2's incentive to exaggerate, the partition in binary communication must be that $\{\theta^1\}, \{\theta^2, \theta^3\}$. Using (8), for any n, θ^2 will not deviate from this partitional strategy if

$$\begin{aligned}\bar{\mu}_2^{\text{bin}}(2) &= \frac{\alpha^0}{\alpha^0+1}\mu + \frac{1}{\alpha^0+1}\bar{\theta}(2) = \frac{\alpha^0}{\alpha^0+1}\underbrace{\frac{7}{16}}_{\mu} + \frac{1}{\alpha^0+1}\left(\underbrace{\frac{5}{8}}_{p^2}\underbrace{\frac{1}{2}}_{\theta^2} + \underbrace{\frac{1}{8}}_{p^3}\underbrace{1}_{\theta^3}\right)\frac{8}{6} \\ &= \frac{\alpha^0}{\alpha^0+1}\frac{7}{16} + \frac{1}{\alpha^0+1}\frac{7}{12} \leq \frac{1}{2}, \end{aligned} \tag{14}$$

which holds for $\alpha^0 \geq 4/3$. In other words, as long as (14) holds, there exists a binary equilibrium for any n. For Table 1 both $\alpha^0 = 7/5$ and $\alpha^0 = 4$ satisfy this condition.

Next, let us consider the condition under which the perfect communication equilibrium exists. Since $\mu < \theta^2$ we can focus on when the middle type mimics the high type θ^3 in a candidate perfect communication equilibrium. Again using (8) we can see that, since $\bar{\theta}(1) = \theta^1$, $\bar{\theta}(2) = \theta^2$, $\bar{\theta}(3) = \theta^3$ under perfect communication, if the middle type reveals truthfully he induces

$$\bar{\mu}_2^{\text{perf}}(2) = \frac{\alpha^0}{\alpha^0+1}\frac{7}{16} + \frac{1}{\alpha^0+1}\frac{1}{2}.$$

If the middle type mimics θ^3 then he induces

$$\bar{\mu}_2^{\text{perf}}(3) = \underbrace{\frac{\alpha^0}{\alpha^0+1}\frac{7}{16} + \frac{1}{\alpha^0+1}\frac{1}{2}}_{\bar{\mu}_2(2)} - \frac{1}{\alpha^0+n}\left(\frac{1}{2} - 1\right).$$

The middle type does not deviate if revealing his type induces the expected policy closer to his ideal i.e., $\left|1/2 - \bar{\mu}_2^{\text{perf}}(2)\right| \leq \left|1/2 - \bar{\mu}_2^{\text{perf}}(3)\right|$, which (for positive α^0 and n) yields

$$\alpha^0 \leq \frac{1}{2}\left(4 - n + \sqrt{n^2 - 8n + 32}\right). \tag{15}$$

It is easy to check that, for Table 1, $\alpha^0 = 7/5$ and $n \leq 5$ satisfies both (14) and (15), and hence support the perfect communication equilibrium up to $n = 5$. Meanwhile (15) does not hold for any $n \geq 1$ when $\alpha^0 = 4$, so that perfect communication cannot be an equilibrium.

The "social welfare" on Table 1 was calculated as follows. For the fully revealing equilibrium, the decision maker's policy conditional on the messages and the prior can be simplified to the first order condition with respect to

$$-\frac{\alpha^0 p^1 + x^1}{\alpha^0+n}(y-0)^2 - \frac{\alpha^0 p^2}{\alpha^0+n}(y-1/2)^2 - \frac{\alpha^0 p^3 + x^3}{\alpha^0+n}(y-1)^2, \tag{16}$$

where x^i denotes the number of sampled agents whose type is θ^i. Note that thanks to the quadratic payoffs (16) also represents the expected payoff of the decision maker conditional on the binary messages. Therefore, by substituting the optimal policy y^* into (16) we obtain the expected payoff conditional on the binary messages. The distribution of types

$\alpha^0 = 1$

n	perfect	binary
0	-0.0898	
7	$*-0.0539$	-0.0643
8	N/A	-0.0633
...		
27	N/A	-0.0591
28	N/A	N/A
$\to\infty$	N/A	N/A

$\alpha^0 = 3/2$

n	perfect	binary
0	-0.0898	
1	$*-0.0755$	-0.0796
2	$*-0.0693$	-0.0753
3	$*-0.0659$	-0.0728
4	$*-0.0637$	-0.0713
5	N/A	-0.0702
$\to\infty$	N/A	$*-0.0643$

$\alpha^0 = 3$

n	perfect	binary
0	-0.0898	
1	$*-0.0842$	-0.0858
2	N/A	-0.0835
3	N/A	-0.0818
4	N/A	-0.0807
5	N/A	-0.0799
$\to\infty$	N/A	$*-0.0739$

Table 2: Decision maker's expected payoff (= "social welfare") when $\theta^1 = 0, \theta^2 = 1/2, \theta^3 = 0, p^1 = 1/8, p^2 = 5/8, p^3 = 2/8$. An asterisk(*) denotes social welfare in neologism proof equilibrium.

is described by the multivariate Pólya distribution. Note that $x^3 = n - x^1 - x^2$. To obtain the numbers on the Tables 1 and 2 we calculated

$$\sum_{x^1=0}^{n}\sum_{x^2=0}^{n-x^1} \frac{\Gamma(\alpha^0)}{\Gamma(\alpha^0+n)} \frac{n!}{x^1!x^2!(n-x^1-x^2)!} \frac{\Gamma(\alpha^0p^1+x^1)\Gamma(\alpha^0p^2+x^2)\Gamma(\alpha^0p^3+n-x^1-x^2)}{\Gamma(\alpha^0p^1)\Gamma(\alpha^0p^2)\Gamma(\alpha^0p^3)} \times u^F(n,x^1,x^2;p,\alpha), \tag{17}$$

where $u^F(n,x^1,x^2,p,\alpha)$ is the maximized expression of (16).

For the binary communication equilibrium the decision maker's payoff is given by

$$-\frac{\alpha^0p^1+x^1}{\alpha^0+n}(y-0)^2 - \frac{\alpha^0p^2+\alpha^0p^3+x^2+x^3}{\alpha^0+n}\frac{p^2}{p^2+p^3}(y-1/2)^2 \\ -\frac{\alpha^0p^2+\alpha^0p^3+x^2+x^3}{\alpha^0+n}\frac{p^3}{p^2+p^3}(y-1)^2. \tag{18}$$

Note that in this case the decision maker cannot distinguish between x^2 and x^3 when deciding the policy. The social welfare is obtained by calculating (17) by replacing $u^F(n,x^1,x^2,p,\alpha)$ with the maximized expression for (18).

7.1 Neologism Proofness

It is clear that by construction, the fully revealing equilibrium is neologism proof if each of the three types *strictly* prefers to play his equilibrium strategy, because no type has incentive to mimic any other type.

To see whether the binary equilibrium $\{\theta^1\}, \{\theta^2, \theta^3\}$ is neologism proof, notice that a sampled agent whose type is θ^3 always wishes to separate himself from θ^2. Therefore, for the binary equilibrium to be neologism proof, i) θ^2 must not have a credible neologism;

and ii) θ^3 must not have a credible neologism, both with respect to the binary equilibrium in question.

Let us assume (14) holds, so that the binary equilibrium exists for any n. It is easy to see that θ^2 does not have a credible neologism since

$$\left|\frac{1}{2}-\underbrace{\left(\frac{\alpha^0}{\alpha^0+1}\frac{7}{16}+\frac{1}{\alpha^0+1}\frac{7}{12}\right)}_{\bar{\mu}_2^{\text{bin}}(2)}\right| < \left|\frac{1}{2}-\underbrace{\left(\underbrace{\frac{\alpha^0}{\alpha^0+1}\frac{7}{16}+\frac{1}{\alpha^0+1}\frac{7}{12}}_{\bar{\mu}_2^{\text{bin}}(2)}-\frac{1}{\alpha^0+n}\left(\frac{7}{12}-\frac{1}{2}\right)\right)}_{\text{expected action when using the neologism, from (9)}}\right|,$$

which holds for any n if $\alpha^0 \geq 4/3$ (recall that $\bar{\mu}_2^{\text{bin}}(2) = 1/2$ when $\alpha^0 = 4/3$). In other words, no types wishes to convince the decision maker that their type is θ^2.

Similarly, θ^3 does not have a credible neologism if θ^2 is better off by convincing the decision maker that he is θ^3, namely

$$\left|\frac{1}{2}-\underbrace{\left(\frac{\alpha^0}{\alpha^0+1}\frac{7}{16}+\frac{1}{\alpha^0+1}\frac{7}{12}\right)}_{\bar{\mu}_2^{\text{bin}}(2)}\right| > \left|\frac{1}{2}-\left(\underbrace{\frac{\alpha^0}{\alpha^0+1}\frac{7}{16}+\frac{1}{\alpha^0+1}\frac{7}{12}}_{\bar{\mu}_2^{\text{bin}}(2)}-\frac{1}{\alpha^0+n}\left(\frac{7}{12}-1\right)\right)\right|, \tag{19}$$

which for $\alpha^0 > 4/3$ gives

$$n > \frac{-3\alpha^2+14\alpha+10}{3\alpha-4}. \tag{20}$$

Therefore, the neologism proofness for a particular pair of α^0 and n can be checked by looking at (20). If $\alpha^0 = 4/3$ then (19) can never be satisfied, which implies there is no neologism binary equilibrium for any n. This is because $\theta^2 = 1/2$ induces the deal expected action $1/2$ for any n in the equilibrium and hence has no incentive to deviate.

References

[1] Austen-Smith, David (1993) "Interested Experts and Policy Advice: Multiple Referrals under Open Rule." *Games and Economic Behavior* 5, 3-43.

[2] Baliga, Sandeep, Corchon, Luis C., and Sjöström, Thomas (1997) "The Theory of Implementation When the Planner Is a Player." *Journal of Economic Theory* 77, 15-33.

[3] Battaglini, Marco (2002) "Multiple Referrals and Multidimensional Cheap Talk." *Econometrica* 70, 1379-1401.

[4] ——— (2004) "Policy Advice with Imperfectly Informed Experts. *Advances in Theoretical Economics* 4, No.1 Article 1.

[5] Crawford, Vincent P., and Sobel, Joel (1982) "Strategic Information Transmission." *Econometrica* 50, 1431-1451.

[6] DeGroot, Morris H. (1970) *Optimal Statistical Decisions*, McGraw-Hill.

[7] Dickey, James M., Jiang, Jhy-Ming and Kadane, Joseph B. "Bayesian Methods for Censored Categorical Data." *Journal of American Statistical Association* 82, 773-781.

[8] Farrell, Joseph (1993) "Meaning and Credibility in Cheap-Talk Games." *Games and Economic Behavior* 5, 514-531.

[9] Feddersen, Timothy, and Pesendorfer, Wolfgang (1997) "Voting Behavior and Information Aggregation in Elections with Private Information." *Econometrica* 65, 1029-1058.

[10] ——— and ——— (1998) "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting." *American Political Science Review* 65, 23-35.

[11] Gerardi, Dino and Yariv, Leeat (2008) "Information Acquisition in Committees." *Games and Economic Behavior* 62, 436-459.

[12] Gershkov, Alex, and Szentes, Balázs (2009) "Optimal Voting Schemes with Costly Information Acquisition." *Journal of Economic Theory* 144, 36-68.

[13] Gilligan, Thomas W., and Krehbiel, Keith (1987) "Collective Decision-Making and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics and Organization* 3, 287-335.

[14] ——— and ——— (1989) "Asymmetric Information and Legislative Rules with a Heterogeneous Committee." *American Journal of Political Science* 33, 459-90.

[15] Goeree, Jacob and Großer, Jens (2007) "Welfare Reducing Polls." *Economic Theory* 31, 51-68.

[16] Kawamura, Kohei (2011) "A Model of Public Consultation: Why is Binary Communication So Common?." *Economic Journal* 121, 819-842.

[17] Krishna, Vijay, and Morgan, John (2001) "A Model of Expertise." *Quarterly Journal of Economics* 116, 747-775.

[18] Morgan, John, and Stocken, Philip C. (2008) "Information Aggregation in Polls." *American Economic Review* 98, 864-896.

[19] Mukhopadhaya, Kaushik (2003) "Jury Size and the Free Rider Problem." *Journal of Law, Economics and Organization* 19, 24-44.

[20] Ottaviani, Marco, and Sørensen, Peter N. (2001) "Information Aggregation in Debate: Who Should Speak First?" *Journal of Public Economics* 81, 393–421.

[21] Persico, Nicola (2004) "Committee Design with Endogenous Information." *Review of Economic Studies* 71, 165-191.

[22] Rothschild, Michael (1974) "Searching for the Lowest Price when the Distribution of Prices is Unknown." *Journal of Political Economy* 82, 689-711.

[23] Talmain, Gabriel (1992) "Search from an Unknown Distribution: An Explicit Solution." *Journal of Economic Theory* 57, 141-157.

[24] Wolinsky, Asher (2002) "Eliciting Information from Multiple Experts." *Games and Economic Behavior* 41, 141-160.