

# Eliciting Information from a Large Population

Kohei Kawamura<sup>\*†</sup>

School of Economics, University of Edinburgh

This Version: February 2013

Published in *Journal of Public Economics*, 103, 44-54, 2013.

## Abstract

This paper studies information transmission in social surveys where a welfare maximizing decision maker communicates with a random sample of individuals from a large population who have heterogeneous preferences. The population distribution of preferences is unknown and has to be estimated, based on answers from the respondents. The decision maker cannot identify the true distribution of preferences even if the sample size becomes arbitrarily large, since the respondents have incentive to "exaggerate" their preferences especially as the sample size becomes larger and each respondent has weaker influence on the decision. The quality of communication with each respondent 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 the decision maker may prefer to sample a *smaller* number of individuals when the prior is *weaker*.

---

<sup>\*</sup>I am grateful to Timothy Feddersen, Leonardo Felli, Navin Kartik, Gary Koop, David Myatt, Hikaru Ogawa, József Sákovics, Larry Samuelson, Dan Silverman, anonymous referees, and seminar participants at Aberdeen, Barcelona (UB), Edinburgh, GRIPS, Nagoya, Waseda, and WZB Berlin for helpful comments and discussions. All errors are my own.

<sup>†</sup>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, the press, and researchers 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. Indeed, policy makers' reliance on opinion polls is well documented,<sup>1</sup> and the design and analysis of surveys are central concerns in marketing research.

It is widely recognized, however, that data collected through surveys may not reflect respondents' true beliefs or opinions. In particular, a disproportionately high proportion of respondents to social surveys tend to choose extreme answers, such as endpoints (e.g. 1 and 5, for a five-point item) on a rating scale. This strongly suggests the presence of a systematic bias in answering survey questions, and indicates that even in the absence of (or after correcting for) selection bias, policy makers and researchers may still have to take into account the possibility of misreporting, when estimating the "true" population distribution.<sup>2</sup>

It should also be noted that the way questions are constructed can also affect the quality of information from respondents. For instance, in opinion polls on current or proposed government policies, questions are asked often in a simple binary form "agree or disagree", even when the preference intensity (how much they agree or disagree) varies across individuals and such information can be of use for policy making. As of 2013 the UK government has held eleven (consultative) referendums, the first in 1973, all of which asked "yes or no" questions.<sup>3</sup> Interestingly, nine of the referendums were about devolution from the central government to local authorities, for which people would have had a wide range of preferences as to how much power should be devolved. Indeed, the actual policy space is not binary either, as the central government ultimately determines the precise degree of devolution, which can also be affected by the margin of votes. Asking simple "yes or no" questions seems to severely limit the information elicited from the public.<sup>4</sup>

---

<sup>1</sup>See e.g. Shapiro (2011) for a survey of the political science literature on the impact of polls on public policies.

<sup>2</sup>For example, online product reviews have disproportionately high numbers of five and one star ratings (Hu et al., 2009). When asked the importance of the referendum on the independence of Scotland from the United Kingdom on the scale of 1 ("should never be done") to 10 ("very important"), 20% (second highest fraction) and 22% (highest fraction) of the sampled respondents gave 1 and 10, respectively (BBC, 2011). See also e.g., Greenleaf (1992), Berinsky (2004), and De Jong et al. (2008) for evidence and discussions on such *extreme response bias*.

<sup>3</sup>Parliamentary sovereignty in the UK implies no national referendums can be binding. Thus they are thought of as information elicitation from the public to inform the central legislature.

<sup>4</sup>For example, the Welsh devolution referendum in 2011 asked the binary question "Do you want the (Welsh) Assembly now to be able to make laws on all matters in the 20 subject areas it has powers for?".

In contrast, small-scale field surveys or consultations with community representatives can be thought of as attempts to obtain more elaborate information from a limited number of respondents. How would the size of a survey affect the incidence of strategic misreporting? If respondents answer strategically and not necessarily honestly, what are the effective ways to ask survey questions? Does a larger sample always lead to better estimation of the underlying population distribution?

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 (or a researcher who informs the decision maker) and a continuum of individuals with heterogeneous preferences. Specifically, we examine how cheap talk communication between the decision maker and randomly sampled individuals changes according to the sample size and the quality of the prior belief about the preference distribution of the population.

One of our main findings is the trade-off between quality of communication and sample size. Needless to say, if every respondent 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 respondent 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 respondent diminishes as the sample size becomes larger. As a result, the population distribution of preferences cannot be inferred precisely, even if the sample size is arbitrarily large. Meanwhile, we also show that some information can be transmitted regardless of sample size, since binary communication leaves no room for exaggeration and thus is informative, although each piece of information obtained from respondents is inevitably coarser than in more detailed communication.

Another related finding, which is perhaps more interesting, is that the decision maker may be better off with sampling a *smaller* number of individuals when the prior belief on the population preferences is *weaker*. This somewhat counterintuitive result comes from the fact that in communication with sampled individuals, each respondent plays not only against the other respondents but also against the decision maker's belief. 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 individuals. 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

---

In theory, the referendum could have asked precisely which matters and which subject areas (including all or none) they wanted legislation power for.

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 respondent has little influence on the decision maker's estimation of the population distribution even if the sample size is small, because a strong prior means that the decision maker's belief is hardly affected by the result of the survey. Consequently, each respondent 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 of binary communication even if, for example, only one respondent is sampled. Insofar as the best available communication is binary, the decision maker can better estimate the population distribution as the sample size becomes larger and thus the optimal sample size is unbounded. This is in sharp contrast to the case where the prior is weak and the optimal sample size is bounded.

The intuition developed in this paper sheds light on the strategic link between the size of a survey and the quality of respondents' answers, which can potentially be of practical use. It indicates that, given the number of choices for questions in a survey, the larger the number of respondents is, the more extreme responses we expect see and they have to be "discounted" for reliable estimation of the population distribution. Asking binary questions in a large survey has the advantage that no correction is required for the incentive to exaggerate, while by design the information on the intensity of the respondents' preferences/opinions is lost. A small survey, where respondents may have less incentive to exaggerate, can outperform a very large one in the estimation of the population distribution.

The feature that binary communication loses information about individuals' preference intensity is studied by Casella and Gelman (2008) for the design of referendums. In their model the binary structure (voting in favour or against a proposal) is exogenous and the decision maker is committed to a majority rule. In our model the decision maker best responds to the communication outcome, as in opinion polls or non-binding referendums. Moreover, binary communication in such large scale communication, which prevails in reality, is endogenously derived in our model. The design problem for the decision maker in the present paper is the size of a survey, which could range from a consultation with a small number of individuals to a large scale opinion poll or non-binding referendum.

A recent paper by Morgan and Stocken (2008) studies information aggregation with cheap talk communication.<sup>5</sup> Similarly to ours, they consider a model with a decision maker and a continuum of individuals with heterogeneous preferences. There are some

---

<sup>5</sup>See e.g., Feddersen and Pesendorfer (1997, 1998) and Goeree and Großer (2007) for information aggregation in the context of voting.

important differences. First, Morgan and Stocken (2008) assume that the message space is binary, so that they do not analyze how the number of messages in equilibrium changes due to the respondents' strategic incentive. Second, since they assume that the distribution of biases of the individuals is known and aggregate uncertainty is only about the distribution of binary signals they receive, information typically aggregates as the size of a poll becomes arbitrarily large. This implies that if polling is costless, the decision maker always prefers to poll an arbitrarily large number of individuals.<sup>6</sup> In our model, on the other hand, information does not aggregate due to the complexity of the underlying state and the limited informativeness of equilibrium communication. 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 individuals 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 belief, 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 affects information transmission from them. He shows that the most informative equilibrium becomes less precise but converges to binary communication as the number of interested parties increases since, as in the present paper, binary messages do not allow exaggeration. However, while Kawamura (2011) offers an insight into why large-scale polls often use binary questions, he assumes that the decision maker communicates with all individuals affected by the decision, which implies the model does not capture important aspects of poll design, namely how large the sample for a poll should be and how the size of a poll affects the informativeness of each response. In contrast, the present paper assumes that the population size is infinite throughout, but the number of randomly sampled individuals the decision maker communicates with is a choice variable. This enables us to study the trade-off between quality and quantity of communication in polls, in relation to the possibility of misreporting and optimal sample size.

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 (Mukhopadhyaya, 2003; Persico, 2004; Gerardi and Yariv, 2008; Gershkov and Szentes, 2009; Koriyama 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 represents their heterogeneous preferences.

---

<sup>6</sup>Also, in Morgan and Stocken (2008) the individuals have both common (ideological) interest and private interest, while we focus the heterogeneity in their private interest.

Other papers that study communication with multiple (mostly two) informed parties include Gilligan and Krehbiel (1987, 1989), Austen-Smith (1993), Krishna and Morgan (2001), Ottaviani and Sørensen (2001), Wolinsky (2002), Battaglini (2002, 2004), Le Quement (2009), and Ishida and Shimizu (2012). Those papers assume that the number of agents who communicate with the decision maker is either fixed, or up to two.

Misreporting in surveys has attracted considerable attention in social studies including social psychology, marketing science, and political science, to name a few. Outside of the economics literature, exaggeration by respondents is often attributed to psychological, cultural and cognitive factors (e.g. Greenleaf, 1992; Berinsky, 2004; and De Jong et al., 2008). In contrast, the present paper offers a model with strategic misreporting, which may be particularly relevant to situations where the result of a survey is likely to influence public policy that in turn affects the welfare of a large number of individuals in a substantive manner.

To our knowledge, this paper offers a first attempt to incorporate complex population uncertainty, which is certainly of interest in public decision making, into a strategic setting. Estimation of distributions is known to be computationally intensive, and in order to keep tractability we introduce the Dirichlet distribution, which is a multinomial extension of the beta distribution.<sup>7</sup> The Dirichlet allows us to explicitly compute posteriors for a rich message structure (including partial pooling of types) and we can easily parametrize the strength of the prior belief/knowledge, which captures how much the decision maker (such as government) is informed about the population before she engages in communication.

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

---

<sup>7</sup>See DeGroot (1970) pp.49-51, 174-175. We also use formulas from Dickey et al. (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).

maximize utilitarian "social welfare"

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

where  $q^i \geq 0$  denotes the proportion of individuals whose type is  $\theta^i$ , such that  $\sum_{i=1}^H q^i = 1$ . Since  $y$  affects all the individuals 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 individuals' (1).

The decision maker does not observe the frequency vector  $q \equiv (q^1, \dots, q^H)$  or any individuals' types directly. Each individual  $a$  learns only his own type  $\theta_a$ . Therefore, the decision maker communicates with randomly sampled  $n$  individuals to estimate  $q$ . In particular, the selected individuals independently send cheap talk messages to the decision maker. In other words, communication is assumed to be completely costless and payoff irrelevant. We allow  $n$  to be any finite number, which can be arbitrarily large.<sup>8</sup> Also,  $n$  is assumed to be common knowledge, which implies that the sampled individuals (respondents) know the size of the survey.<sup>9</sup>

We assume that the frequency vector  $q$  follows the Dirichlet distribution with parameters  $\alpha \equiv (\alpha^1, \alpha^2, \dots, \alpha^H) = (\alpha^0 p^1, \alpha^0 p^2, \dots, \alpha^0 p^H)$  such that  $\alpha^0 > 0$ ,  $p^i > 0$  for all  $i = 1, \dots, H$ , 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

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

where  $\Gamma(\cdot)$  denotes the gamma function

$$\Gamma(\zeta) \equiv \int_0^\infty t^{\zeta-1} e^{-t} dt \text{ for } \zeta > 0.$$

The density function may look complex but a convenient and intuitive 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  $i$ th type is

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

Therefore we can see  $p \equiv (p^1, p^2, \dots, p^H)$  as the expected prior population distribution of preferences. The decision maker and all individuals share the same prior, before the

---

<sup>8</sup>We rule out the case where  $n$  is infinitely large, so that each respondent's message has some (however small) influence on the decision maker's belief and policy.

<sup>9</sup>In practice, respondents should be able to have a good idea about the size of a survey, from the way the survey is conducted and how it is publicized. For instance, detailed in-person interviews would be adopted for a relatively small sample, while phone polls or online questionnaire would typically be used for a larger sample.

individuals learn their types. In our common prior Bayesian framework, each individual 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 respondents. Let the prior expected mean of the individuals' types be

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

For expositional convenience we assume  $\mu \neq \theta^i$  for all  $i = 1, \dots, H$ . In other words, we rule out the non-generic case where the prior mean coincides exactly with one of the types. Suppose that every respondent reveals their type truthfully. Let  $x = (x^1, \dots, x^H)$  be the count vector where  $x^i$  denote the number of individuals whose type is  $\theta^i$ , out of  $n$  respondents. 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

$$E[q^i | x^i] = \frac{\alpha^0 p^i + x^i}{\alpha^0 + n}. \quad (3)$$

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 individual  $x^{-i}$ .

From each individual'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

$$\begin{cases} E[q^i | \theta_a = \theta^i] = \frac{\alpha^0 p^i + 1}{\alpha^0 + 1} \\ E[q^{-i} | \theta_a = \theta^i] = \frac{\alpha^0 p^{-i}}{\alpha^0 + 1} \end{cases} \quad (4)$$

That is, (4) describes the expected posterior distribution of the population preferences from the viewpoint of an individual whose type is  $\theta^i$ . Note that each individual 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,  $\alpha^0$  can be interpreted as the "strength" of the prior belief. That is, from (3) the "sensitivity" of the posterior with respect to the count vector of the respondents' types

$$\frac{\Delta E[q^i | x^i]}{\Delta x^i} \quad (5)$$

is strictly decreasing in  $\alpha^0$ . This implies that the prior is influenced by the sample less (and hence the prior belief is "stronger"), as  $\alpha^0$  becomes larger.<sup>10</sup> In addition,  $\alpha^0$  can also be

---

<sup>10</sup>We can also see  $\alpha^0$  as inversely related to the informativeness (strength) of a given set of data.

seen as the level of ex ante aggregate uncertainty, conditional on prior common knowledge about the population distribution. When  $\alpha^0$  is high, the realized population distribution is likely to be similar to the prior. For example, if  $\alpha^0 \rightarrow \infty$  then  $E[q^i | x^i] \rightarrow E[q^i] = p^i$  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  $\alpha^0$  is a finite number, the realized distribution may well be different from the prior and there is uncertainty in the population distribution of preferences.

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

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

In other words, the lower  $\alpha^0$  is, the more likely the others are of his type. In particular, if  $\alpha^0 \rightarrow 0$  we have  $E[q^i | \theta_a = \theta^i] \rightarrow 1$ , which means the other individuals' 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. The link between the weakness of the prior as measured in (5) and correlation takes an extreme form under the Dirichlet assumption especially for  $\alpha^0$  close to 0. As we will see later, what is necessary for the intuition behind our results is that, an additional observation of a particular type skews the posterior towards that type, and the magnitude of the additional skew is parametrized monotonically by a single variable ( $\alpha^0$  under the Dirichlet). This feature requires some form of correlation of types but not necessarily the Dirichlet (apart from its tractability). For example, the same intuition would hold if from an individual's viewpoint the types of the others are correlated not only to his own type but also types close to his.<sup>11</sup>

The timing of the game is as follows:

1. All individuals and the decision maker are endowed with a common prior on the preference distribution;
2. individuals privately learn their types;

---

<sup>11</sup>The negative association between the strength of the prior and correlation follows directly from the Dirichlet assumption. 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 particular, from each individual's viewpoint the correlation is with respect to his own type only, and not to any other types. 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.

3. The decision maker randomly samples  $n$  individuals 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 individuals 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 individual has only one (his own type).

### 3 Equilibrium

Throughout this paper we focus on symmetric partitional strategies of the individuals, in which there are  $K$  non-overlapping groups, each of which consists of one or more consecutive type indices. Naturally we have  $K \leq H$ , that is, the number of groups is weakly smaller than the number of types. Any respondents in the same type group induce (from their viewpoint) the same distribution of policy by the decision maker, and without loss of generality we assume that all respondents in the same group send an identical message.<sup>12</sup> As in any cheap talk models, there may be multiple equilibria in our model, and in particular for any parameter values there exists an uninformative ("babbling") equilibrium where all respondents send uninformative messages and the decision maker chooses her policy based only on her prior. However, the decision maker is strictly better off in an informative equilibrium as she can use additional information from respondents to maximize her conditional expected payoff. This also implies that all individuals are ex ante better off in an informative equilibrium than in the uninformative equilibrium, since ex ante they share common interest with the decision maker. In what follows we will derive informative equilibria of the game.

Let  $G^k$  be the set of type indices in the  $k$ th group from the left 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:  $G^1 = \{1, 2, \dots, H\}$ . Let  $z \equiv (z^1, \dots, z^k, \dots, z^K)$  be the count vector of messages from the respondents in each

---

<sup>12</sup>Respondents 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 respondents in the same group send an identical message, in the sense that the same combination of the respondents' types results in the same policy.

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:

$$\bar{y}(z) = \sum_{i=1}^H E[q^i | z] \times \theta^i, \quad (6)$$

where  $E[q^i | z]$  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  $k$ th 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 respondents are of type  $\theta^i$ . The expected frequency of each type  $\theta^i$ , conditional on the count vector of messages  $z$  is given by

$$E[q^i | z] = \frac{\alpha^0 \sum_{l \in G(i)} p^l + z(i)}{\underbrace{\alpha^0 + n}_{\text{expected frequency of } G(i)}} \frac{p^i}{\underbrace{\sum_{l \in G(i)} p^l}_{\text{weight within } G(i)}}, \quad (7)$$

where  $z(i)$  denotes the number of respondents in  $G(i)$ , i.e.  $z(i) \equiv \sum_{l \in G(i)} x^l$ .<sup>13</sup> Since the decision maker does not observe  $x^i$  directly for group  $G(i)$  that contains two or more types, the estimation of  $q^i$  is more complex than in (3), where  $x^i$  is known. We can interpret the construction of (7) in the following two steps: i) we first calculate the expected frequency of  $G(i)$  respect to the other groups; and ii) "allocate" the expected frequency of the group according to the relative prior frequency (weight) of  $\theta^i$  within the group. The first step is analogous to (3) but here the difference is that the numerator involves the sum of prior frequencies and the number of messages for the group, rather than those of a specific type.

From each respondent's viewpoint, his message induces a distribution of the decision maker's policy, which is influenced also by the other respondents' messages. Note from (6) and (7) that, for any possible count vector of the other respondents, the difference in the induced policy by sending two different messages, respectively for  $G^j$  (thereby increasing  $z^j$  by one) and  $G^k$  (increasing  $z^k$  by one), is given by

$$\sum_{i \in G^j} \frac{1}{\alpha^0 + n} \frac{p^i}{\sum_{l \in G^j} p^l} \theta^i - \sum_{i \in G^k} \frac{1}{\alpha^0 + n} \frac{p^i}{\sum_{l \in G^k} p^l} \theta^i, \quad (8)$$

which is independent from the messages from the other respondents. This implies that from each respondent's viewpoint his message influences the expectation but not the variance of the policy. Therefore, as we assume quadratic payoffs, we can focus on the expected policy induced by each message when we consider a respondent's strategy.

Every individual updates his own belief on the population distribution, according to his own type. Given that the individual's type is  $\theta^i$ , the types of all the other individuals

---

<sup>13</sup>See Dickey et al. (1987), pp777-780.

are Dirichlet distributed with parameters  $\alpha' = (\alpha^1, \dots, \alpha^i + 1, \dots, \alpha^H)$ . 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  $\theta^{-i}$  denotes a type other than  $\theta^i$ .

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):

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

where  $x^i$  is the number of respondents whose type is  $\theta^i$ .

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

$$u(\theta^i, m(i)) = - \sum_{x^1=0}^n \sum_{x^2=0}^{n-x^1} \dots \sum_{x^H=0}^{n-x^1-\dots-x^{H-1}} \Pr(x | \alpha') \left( \theta^i - \sum_{t=1}^H E[q^t | z^1, \dots, z(i) + 1, \dots, z^K] \times \theta^t \right)^2,$$

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

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

Note that the expected payoff function  $u(\theta^i, \cdot)$  already incorporates (6), the decision maker's best response to the messages from the respondents given that all of them follow the partitional strategy. Hence partitional strategies form a perfect Bayesian equilibrium if, for any  $\theta^i$

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

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

How does communication between the decision maker and the respondents take place in equilibrium? As we will see clearly in the next section, unlike cheap talk models with

continuous types there can be a fully revealing equilibrium when the prior is weak ( $\alpha^0$  low) and the sample size is small. Since in such a case each respondent has relatively large influence on the policy and thus less incentive to exaggerate, the best response is to reveal his type truthfully rather than mimicking another type and thereby shifting the expected policy further away from his ideal.<sup>14</sup> This points to the possibility that we have more informative equilibria with each responder, when a small sample size is combined with a weak prior.

Meanwhile, the following proposition states that, when the prior belief about the population 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 respondents' types are partitioned into only two groups, and any respondent from a group induces the same belief as the other respondents in the group. In other words, the decision maker can correctly infer to which of the two type groups a respondent belongs, but cannot precisely know the respondent's type.

**Proposition 1** If the prior belief about the population distribution of preferences is sufficiently strong ( $\alpha^0$  is sufficiently large), a binary equilibrium in which only two messages are used exists for any sample size  $n$ , whereby all types below the ex ante average type  $\mu$  send one message and those above  $\mu$  send the other. The binary equilibrium is the only informative equilibrium in partitional strategy.

**Proof.** See Appendix I. ■

As we have already suggested in the Introduction, this proposition has a simple intuition. Note that, the prior stronger is, the smaller influence each respondent has on the decision maker's belief and hence her policy, regardless of the sample size. Also, the expected prior policy from each respondent's viewpoint becomes closer to the prior expectation  $\mu$ . Suppose there are three or more groups ( $K \geq 3$ ) partitioning the type space, and consider a middle group which is neither the bottom  $G^1$  or the top  $G^K$ . When  $\alpha^0$  is high, a respondent in such a group whose ideal policy is lower (higher) than  $\mu$  deviates and mimics one in a lower (higher) group, since by doing so he can render the expected policy closer to his ideal. In other words, respondents whose types are above  $\mu$  wish to overstate their types as much as they can, and respondents below  $\mu$  understate their types as much as they can. Binary communication is "robust" to the incentive to exaggerate, because the respondents may not possibly exaggerate their types when they have the choice between two messages (above or below  $\mu$ ).

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

---

<sup>14</sup>If the type space is continuous, even infinitesimally small incentive to exaggerate leads to misreport as a neighbourhood type, so that the fully revealing equilibrium would not exist.

**Proposition 2** For sample size  $n$  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: if there are more than three groups, respondent types in a middle group deviate and mimic one in an extreme group due to their weak influence on the policy.<sup>15</sup> On the other hand, 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  $\alpha^0$  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 respondent who finds himself having this type, the expected action (given that the other respondents follow the binary partitional strategy) will be *lower* than his ideal, since the other respondents 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 respondent 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  $\alpha^0$  and large  $n$  there could be a mixed strategy equilibrium where respondents 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.<sup>16</sup> However, in our framework very small  $\alpha^0$  has a somewhat unrealistic feature that, due to high correlation, a vast majority of the individuals 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 conflict of interest between the decision maker and any individuals. In what follows we focus on  $\alpha^0$  such that a binary equilibrium exists, which implies we expect to see at least some dispersion of realized preferences.

---

<sup>15</sup>Kawamura (2011) has given a related proposition but with a finite number of individuals 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 individuals, 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  $\theta^1, \theta^2, \dots, \theta^H$  even when the sample size is very small or when no respondent turns out to be of certain types.

<sup>16</sup>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.

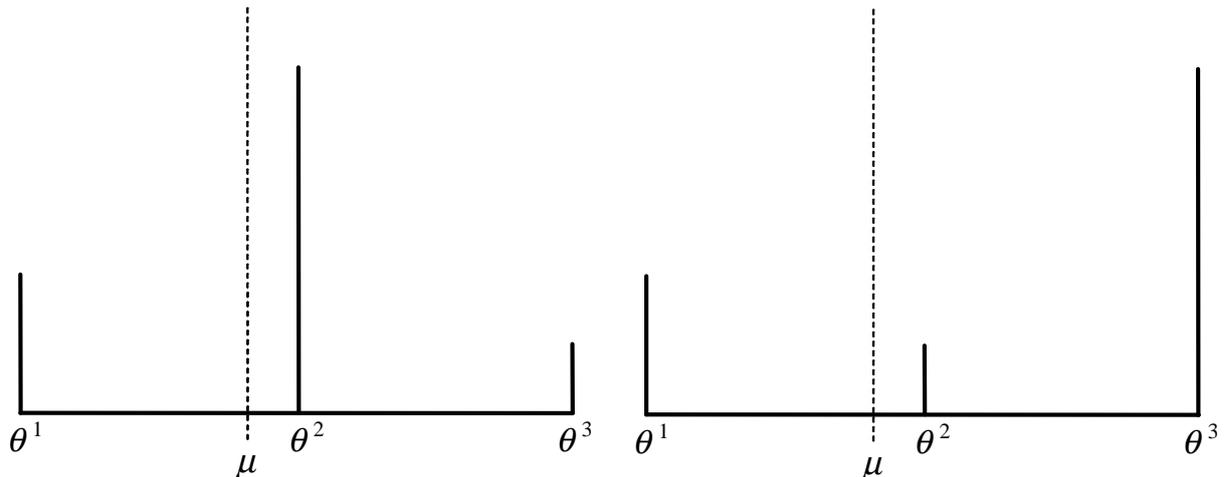


Figure 1: Two observationally equivalent preference distributions for binary data

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 individuals are below and above  $\mu$ , but she can never accurately infer how the types are distributed above  $\mu$ . This makes it impossible for the decision maker to implement the first best policy for any sample size, and suggests the possibility that limiting the sample size may improve the decision maker's estimation, which is the focus of the next section.

Propositions 1 and 2 assume that there does not exist a type that coincides with the prior expected mean  $\mu$ . If such a type exists, this type does not have incentive to exaggerate as his posterior expectation on the other respondents' 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 all individuals whose type coincides with  $\mu$  send a distinct message, there does not exist a fully revealing equilibrium for large  $n$  or  $\alpha^0$ .

So far we have assumed  $H \geq 3$ , so as to capture the complexity of the population distribution of preferences. When the distribution has a simpler form, information does aggregate for large  $n$ , as already found in e.g. Morgan and Stocken (2008):

**Remark 1** If there are only two types ( $H = 2$ ), then the binary communication (full revelation) equilibrium exists for any  $\alpha^0$  and  $n$ . The information aggregates for  $n$  arbitrarily large.

With two types, each type renders the decision maker's action closer to their ideal only by revealing truthfully, since mimicking the other type merely shifts it away from his ideal policy. That is, there is neither any room for exaggeration nor incentive to misreport, and hence respondents reveal their types truthfully. Remark 1 highlights the importance of

the complexity (specifically, the presence of varying preference intensity) in the underlying preference distribution for our analysis.

What if the decision maker's action  $y$  is binary for exogenous reasons? This may apply to, for instance, the ratification of a treaty or the choice of two candidates for a particular position who are perfectly committed to fixed policies. If the preferences are also binary ( $H = 2$ ), then there is no strategic incentive to misreport and the decision maker knows the true distribution. However, if  $H \geq 3$ , while the decision maker wants to implement the action closer to the average type of the population, it may not be correctly estimated even with an arbitrarily large sample, for the same reason as in the case where the policy space is continuous. To see this, suppose the prior distribution is such that the decision maker's optimal choice from the two possible actions  $y \in \{y^0, y^1\} \subset \mathbb{R}$  changes according to the true population distribution. Then if the decision maker takes the respondents' messages at face value, every non-extreme respondent has incentive to mimic the extreme type whose favoured policy is the same as his. This does not depend on the sample size, because even if a non-extreme respondent has large influence on policy (because of small  $\alpha^0$  or  $n$ ), he does not gain by revealing truthfully since the policy cannot be "fine-tuned" in response to his message.<sup>17</sup> At the same time, the binary equilibrium exists as it is consistent with the respondents' incentive. The above discussion is summarized as follows:

**Remark 2** If the policy space is binary and there are three or more types ( $H \geq 3$ ), then the binary equilibrium is the only informative partitional equilibrium for any sample size.

Remarks 1 and 2 indicate that both type space and policy space have to be richer than binary for sample size to affect the quality of communication. However, as we have discussed in the Introduction, in many situations of interest the actual policy space is much richer than binary, while it may well be presented as binary in large-scale surveys or referendums. Indeed, in light of Proposition 2 we can understand such binary questions as an equilibrium outcome under a non-binary environment. In the following, we return to  $H \geq 3$  and continuous policy space to study the choice of sample size.

## 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 individual, 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

---

<sup>17</sup>In other words, all he can do is to increase the probability of his preferred policy (between the two) being implemented, by pretending to be an extreme type.

from respondents.

When the underlying uncertainty on the preference distribution is 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 individuals 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 respondent 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 respondents, by assuming that the decision maker can commit to a sample size and the respondents know it. This is an important assumption in our model not least because conditional on a certain number of received messages from the respondents, the decision maker is always tempted to sample more to estimate the population distribution better.<sup>18</sup> Likewise, individuals in the population who are not sampled would always like to communicate and influence the policy in their favour. If the decision maker cannot commit to an announced sample size, the respondents would anticipate that the actual sample size is arbitrarily large, in which case only binary communication is available. 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's expected payoff conditional on the sample size  $n$  and the best response (6) is computed by

$$u^{DM}(n) = - \sum_{x^1=0}^n \sum_{x^2=0}^{n-x^1} \dots \sum_{x^H=0}^{n-x^1, \dots, -x^{H-1}} \underbrace{\Pr(x \mid \alpha)}_{\text{dist. of type counts}} \underbrace{\left( \sum_{i=1}^H E[q^i \mid z] (\bar{y}(z) - \theta^i)^2 \right)}_{\text{exp. payoff conditional on message counts}}, \quad (9)$$

where  $\Pr(x \mid \alpha)$  denotes the multivariate Pólya distribution,  $z = (z^1, \dots, z^K)$  is the count vector of the messages in a partitioned 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

---

<sup>18</sup>Throughout this paper we maintain the assumption that the decision maker does not commit to any *decision rule* that ex ante specifies the policy to be implemented according to the messages sent. This assumption would be appropriate to analyze information transmission through interviews with representatives, opinion polls, and non-binding referendums where the decision maker determines the policy after obtaining information from a population.

$\alpha^0 = 1.5$			$\alpha^0 = 4.5$		
$n$	full	binary	$n$	full	binary
0	-0.0898		0	-0.0898	
1	*-0.0742	-0.0788	1	N/A	-0.0877
2	*-0.0678	-0.0742	2	N/A	*-0.0863
3	*-0.0643	-0.0717	3	N/A	*-0.0852
4	*-0.0621	-0.0702	4	N/A	*-0.0844
5	*-0.0606	-0.0691	5	N/A	*-0.0837
6	N/A	-0.0683	6	N/A	*-0.0832
$\rightarrow \infty$	N/A	*-0.0633	$\rightarrow \infty$	N/A	*-0.0782

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

$\theta^2$ , which is above the expected mean, has incentive to exaggerate and mimic the high type  $\theta^3$ . There are multiple equilibria in this game, and the payoff in the uninformative equilibrium is the same as that of no communication at all  $n = 0$ . The most informative equilibrium in partitional strategy is either fully revealing, in which case each type  $\theta^i$  for  $i = 1, 2, 3$  sends a distinct message; or binary, in which case respondents with  $\theta^1$  and  $\theta^2$  send the same message and those with  $\theta^3$  send a separate message.

Assuming that all respondents play the same equilibrium strategy, we can see that, if  $\alpha^0 = 1.5$  and the most informative equilibrium is chosen, the social welfare is non-monotonic in the sample size since the welfare under full revelation for  $n = 5$  (-0.0606) is higher than the welfare under binary communication for  $n = 6$  (-0.0683). Moreover, the social welfare for  $n = 5$  is higher than for  $n \rightarrow \infty$ , which implies that the optimal sample size is bounded even if sampling itself is completely costless. This is because for  $n \geq 6$  the incentive to exaggerate is so strong for the respondents that, there does not exist an equilibrium where every type reveals truthfully. When the prior is stronger (the expected payoffs are listed for  $\alpha^0 = 4.5$ ), each respondent 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 individuals when the prior is *weaker*. Figure 2 shows how the optimal sample size changes according to  $\alpha^0$ . As we discussed in the previous section, when  $\alpha^0$  is very small the binary equilibrium may not exist. In this example  $\alpha^0 \geq 4/3$  guarantees its existence for any  $n$ . It is easy to check

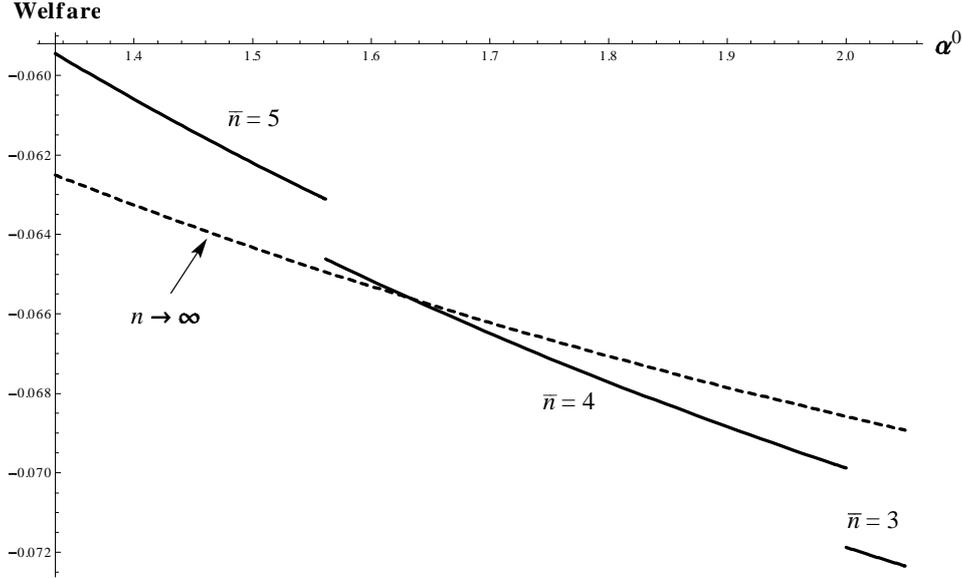


Figure 2: Welfare under full ( $\bar{n}$ ) and binary ( $n \rightarrow \infty$ ) communication

that the largest sample size that supports full revelation, denoted by  $\bar{n}$ , is decreasing in  $\alpha^0$ . Intuitively, as the decision maker's belief is less influenced by the respondents' messages, the incentive to exaggerate becomes stronger, which makes it harder to sustain full revelation. The solid lines in Figure 2 represent the welfare under full revelation with the largest feasible sample size, and the dashed line represents the welfare under binary communication with an arbitrary large sample.<sup>19</sup> We can observe that, for  $\alpha^0 \geq 4/3$ , the optimal sample size is 5 up to  $\alpha^0 \approx 1.562$ , and 4 when  $\alpha^0$  is between 1.562 and 1.633. For  $\alpha^0$  larger than 1.633, although full communication can be supported in an equilibrium for small  $n$ , the optimal sample size is unbounded. Thus the optimal sample size is non-monotonic in the strength of the prior.

Our argument regarding optimal sample size is based on symmetric strategies. In principle, if we allow asymmetric strategies, having a larger sample than the "optimal" one above may never hurt, since those additionally respondents can play the uninformative equilibrium without decreasing the welfare. However, we can interpret our result with asymmetric strategies (or equilibrium), as long as we define optimal sample size to be the smallest  $n$  that yields the highest welfare. This can be justified, for example, if there is a small sampling cost that is increasing in  $n$  and the decision maker knows which communication equilibrium she will play when deciding  $n$ . Alternatively, we can use the uninformative equilibrium for respondents beyond  $n$  as a justification for pre-determined sample size: even

<sup>19</sup>The reason why the welfare is decreasing in  $\alpha^0$  is that the prior expected distribution is relatively dispersed in this setup. In general, if the population is dispersed the welfare tends to be lower, and here higher  $\alpha^0$  means that the realized distribution is indeed more likely to be dispersed. When the prior expected distribution is concentrated, the welfare in an informative equilibrium can be increasing in  $\alpha^0$ .

without commitment to sample size  $n$ , the decision maker can play the most informative equilibrium with  $n$  respondents, and the uninformative equilibrium with the others.

The discussion in this section can be summarized as follows.

**Proposition 3** If the prior belief about the population distribution of preferences is sufficiently strong ( $\alpha^0$  is sufficiently large), the optimal sample size is unbounded. Otherwise, the relationship between social welfare and sample size may be non-monotonic and the optimal sample size may be bounded. The relationship between  $\alpha^0$  and the optimal sample size may be non-monotonic.

The result that the optimal sample size is unbounded when  $\alpha^0$  is sufficiently large follows from Proposition 1: if binary communication is the only informative equilibrium, then a larger sample unambiguously leads to higher welfare.

We have focused on perfect Bayesian equilibria of the game, but in practice, respondents to a survey may not necessarily answer the questions and instead might communicate in their own way. For example, a respondent may send a detailed message about his preference, even when he is asked a "yes or no" question. How would the decision maker react to such an off-the-equilibrium message? As indicated in Table 1, the binary equilibrium is "neologism proof" (Farrell, 1993) with a large sample size and/or  $\alpha^0$ , which means binary communication is not only simple but also robust to potential off-the-equilibrium messages.<sup>20</sup> In other words, any non-binary (perhaps more detailed) off-the-equilibrium message by a respondent cannot be credible and hence must be no more informative than a binary message.

The intuition for the robustness of binary communication can be presented somewhat more precisely as follows. Note that in the example of this section,  $\theta^1$  and  $\theta^2$  are the types of respondents who may wish to send a more "detailed" message because they pool and send the same coarse message in equilibrium. Clearly a respondent whose type is  $\theta^1$  wishes separate himself from  $\theta^2$  by using an off-the-equilibrium message (neologism), because if he successfully convinces the decision maker of his type he can render the policy lower and thus closer to his ideal. However, since  $\theta^2$  is lower than the prior mean  $\mu = 7/16$ , their incentive under large sample size and/or  $\alpha^0$  is also that they want to convince the decision maker that their type is extreme ( $\theta^1$ ) by using a neologism. Insofar as both  $\theta^1$  and  $\theta^2$  in the binary equilibrium want to convince the decision maker that their type is  $\theta^1$ , such a neologism can never be credible (because both types would use it), and indeed no more informative than the original binary communication where  $\theta^1$  and  $\theta^2$  send the same message. This also implies that in order to play the binary equilibrium with a large sample size the decision maker does *not* need commitment to the restricted message space.

---

<sup>20</sup>See Appendix II for a detailed discussion on neologism proofness in the example of this section.

## 5 Concluding Remarks

This paper has studied information transmission in communication with sampled individuals from a large population. Our results shed light on the design and interpretation of small to large scale social surveys and non-binding referendums, which have substantial influence on public decisions in reality. The model developed in this paper offers an insight into why large surveys and referendums ask simple, often binary "yes or no", questions; and why they attract extreme responses systematically when non-binary questions are asked. In particular, we highlight the trade-off between the quality and quantity communication caused by respondents' strategic incentive to misreport. Since a large sample size may diminish the quality of communication with each respondent, the optimal sample size may be bounded, even if communication and information processing are completely costless. We have demonstrated that this is especially 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 individuals 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 individuals, 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 individuals. 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  $\{\theta^1, \theta^2, \dots, \theta^H\}$  into  $J(\leq H)$  disjoint groups. Let  $\theta^{(j,1)}$  be the lowest and  $\theta^{(j,S_j)}$  be the highest type in the  $j$ th group that consists of  $S_j$  types. Let  $\pi : (j, s) \mapsto \{1, 2, \dots, H\}$  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 any partition such that  $J \geq 3$  and  $\theta^{(2,1)} < \mu$  cannot be an equilibrium if  $\alpha^0$  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  $J$ th group.

First, let us denote the expected type of an individual in the  $j$ th 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 individual types. As we have seen in (8) this ensures that the variance of the decision maker's policy from the respondent'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 respondent's ideal policy.

Regardless of the partition of the type space, the ex ante expected type of the individuals from the decision maker's viewpoint is  $\mu$ . 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 respondent has learnt his type, and let us consider from his viewpoint how the other respondents 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 respondents in group  $j$ . We have  $\sum_{j=1}^J z^j = n$ . If all respondents follow the partitional strategy, the decision maker's policy conditional on their messages is given by

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

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

$$\begin{aligned} \bar{\mu}_2(2) &\equiv \sum_{j=1}^J \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)} + E[z^j \mid \theta_i = \theta^{(2,1)}]}{\alpha^0 + n} \bar{\theta}(j) + \frac{1}{\alpha^0 + n} \bar{\theta}(2) \\ &= \sum_{j=1, j \neq 2}^J \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)} + (n-1) \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)}}{\alpha^0 + 1}}{\alpha^0 + n} \bar{\theta}(j) + \frac{\sum_{s=1}^{S_2} \alpha^{(2,s)} + (n-1) \frac{\sum_{s=1}^{S_2} \alpha^{(2,s)} + 1}{\alpha^0 + 1}}{\alpha^0 + n} \bar{\theta}(2) \\ &= \frac{\alpha^0}{\alpha^0 + 1} \mu + \frac{1}{\alpha^0 + 1} \bar{\theta}(2). \end{aligned} \tag{10}$$

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

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

$$\begin{aligned}\bar{\mu}_2(1) &= \frac{\sum_{s=1}^{S_1} \alpha^{(1,s)} + E[z^1 | \theta_i = \theta^{(2,1)}] + 1}{\alpha^0 + n} \bar{\theta}(1) + \sum_{j=2}^J \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)} + E[z^j | \theta_i = \theta^{(2,1)}]}{\alpha^0 + n} \bar{\theta}(j) \\ &= \sum_{j=1}^J \frac{\sum_{s=1}^{S_j} \alpha^{(j,s)}}{\alpha^0 + 1} \bar{\theta}(j) + \frac{1}{\alpha^0 + n} \bar{\theta}(1) + \frac{(n-1)}{(\alpha^0 + n)(\alpha^0 + 1)} \bar{\theta}(2) \\ &= \bar{\mu}_2(2) - \frac{1}{\alpha^0 + n} (\bar{\theta}(2) - \bar{\theta}(1)).\end{aligned}\tag{11}$$

We now observe that for large enough  $\alpha^0$

$$\theta^{(2,1)} < \bar{\mu}_2(1) < \bar{\mu}_2(2),\tag{12}$$

which implies that the decision maker's policy is closer to his ideal when he mimics a respondent in the first group whose expected type is lower than his own. Note that from (6) and (7) the respondent's message does not influence the variance of the decision maker's policy. Thus (12) implies that the expected payoff of a respondent in the second group is higher if he mimics one in the first group.

Similarly, consider a respondent 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 respondent  $\theta^{(J-1, S_{J-1})} > \mu$ , the decision maker's expectation on any other individual is lower than his own type. Therefore, for large enough  $\alpha^0$ , the decision maker's policy is closer to his ideal policy when he mimics a respondent in the  $J$ th group:

$$\theta^{(J-1, S_{J-1})} > \bar{\mu}_{J-1}(J) > \mu_{J-1}(J-1).\tag{13}$$

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

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

$$\theta^1 < \dots < \theta^{(1, S_1)} < \bar{\mu}_1(1) = \frac{\alpha^0}{\alpha^0 + 1} \mu + \frac{1}{\alpha^0 + 1} \bar{\theta}(1)\tag{14}$$

and

$$\bar{\mu}_1(2) = \frac{\alpha^0}{\alpha^0 + 1} \mu + \frac{1}{\alpha^0 + 1} \bar{\theta}(2) < \theta^{(2, 1)} \dots < \theta^H.\tag{15}$$

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

## 6.2 Proposition 2

**Proof.** From (11), (12) and also (13) hold for large enough  $n$ , regardless of  $\alpha^0$ . This rules out any partitional equilibrium with three or more groups.

A binary partition equilibrium does not exist either, if  $n$  is large and  $\alpha^0$  is close to 0. To see this, consider any binary partition where the second group contains two or more types. Then for  $\alpha^0$  close enough to 0

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

Hence from (11) 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 respondent whose type is  $\theta^{(2,1)}$  mimics a respondent 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  $\alpha^0$  is close to 0. This completes the proof of the first part of the proposition.

Meanwhile, if  $\alpha^0$  is not too small, with respect to the binary partition such that  $\theta^{(1,S_1)} < \mu < \theta^{(2,1)}$ , both (14) and (15) hold. Therefore Proposition 1 implies that the partition constitutes an equilibrium for any  $n$ , and it is the only informative partitional equilibrium for large enough  $n$ . ■

## 7 Appendix II (neologism proofness)

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  $\theta^2$ , 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  $\theta^2$ 's incentive to exaggerate, the partition in binary communication must be that  $\{\theta^1\}, \{\theta^2, \theta^3\}$ . Using (10), for any  $n$ ,  $\theta^2$  will not deviate from this partitional strategy if

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

which holds for  $\alpha^0 \geq 4/3$ . In other words, as long as (16) holds, there exists a binary equilibrium for any  $n$ . For Table 1 both  $\alpha^0 = 7/5$  and  $\alpha^0 = 4$  satisfy this condition.

$\alpha^0 = 1$			$\alpha^0 = 2.5$		
$n$	perfect	binary	$n$	perfect	binary
0	-0.0898		0	-0.0898	
7	*-0.0539	-0.0643	1	*-0.0825	-0.0846
8	N/A	-0.0633	2	*-0.0784	-0.0817
...			3	*-0.0758	-0.0799
27	N/A	-0.0591	4	N/A	-0.0786
28	N/A	N/A	5	N/A	-0.0777
$\rightarrow \infty$	N/A	N/A	$\rightarrow \infty$	N/A	*-0.0716

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

Next, let us consider the condition under which the fully revealing equilibrium exists. Since  $\mu < \theta^2$  we can focus on when the middle type mimics the high type  $\theta^3$  in a candidate equilibrium with full revelation. Again using (10) we can see that, since  $\bar{\theta}(1) = \theta^1, \bar{\theta}(2) = \theta^2, \bar{\theta}(3) = \theta^3$  under full revelation, 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  $\theta^3$  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  $\alpha^0$  and  $n$ ) yields

$$\alpha^0 \leq \frac{1}{2} \left( 4 - n + \sqrt{n^2 - 8n + 32} \right) \equiv \bar{\alpha}(n). \quad (17)$$

It is easy to check that, for Table 1,  $\alpha^0 = 1.4$  and  $n \leq 5$  satisfies both (16) and (17), and hence support the fully revealing equilibrium up to  $n = 5$ . Meanwhile if  $\alpha^0 > 4$  (17) does not hold for any  $n \geq 1$ . Thus, for example, when  $\alpha^0 = 4.5$  a fully revealing equilibrium does not exist.

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

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

where  $x^i$  denotes the number of respondents whose type is  $\theta^i$ . Note that thanks to the quadratic payoffs (18) also represents the expected payoff of the decision maker conditional on the binary messages. Therefore, by substituting the optimal policy  $y^*$  into (18) we obtain the expected payoff conditional on the binary messages. The distribution of types 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

$$\sum_{x^1=0}^n \sum_{x^2=0}^{n-x^1} \frac{\Gamma(\alpha^0)}{\Gamma(\alpha^0 + n)} \frac{n!}{x^1!x^2!(n-x^1-x^2)!} \frac{\Gamma(\alpha^0 p^1 + x^1)\Gamma(\alpha^0 p^2 + x^2)\Gamma(\alpha^0 p^3 + n - x^1 - x^2)}{\Gamma(\alpha^0 p^1)\Gamma(\alpha^0 p^2)\Gamma(\alpha^0 p^3)} \times u^F(n, x^1, x^2; p, \alpha), \quad (19)$$

where  $u^F(n, x^1, x^2, p, \alpha)$  is the maximized expression of (18).

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

$$\begin{aligned} & -\frac{\alpha^0 p^1 + x^1}{\alpha^0 + n} (y - 0)^2 - \frac{\alpha^0 p^2 + \alpha^0 p^3 + x^2 + x^3}{\alpha^0 + n} \frac{p^2}{p^2 + p^3} (y - 1/2)^2 \\ & - \frac{\alpha^0 p^2 + \alpha^0 p^3 + x^2 + x^3}{\alpha^0 + n} \frac{p^3}{p^2 + p^3} (y - 1)^2. \end{aligned} \quad (20)$$

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 (19) by replacing  $u^F(n, x^1, x^2, p, \alpha)$  with the maximized expression for (20). For the binary equilibrium with an arbitrarily large sample, we compute the expected payoff when the decision maker has full information about the relative sizes of the first and the second group for any realization of the type distribution.

## 7.1 Neologism Proofness

Here we describe how neologism proofness of an equilibrium in the example can be checked, and also show that if binary equilibrium exists for any  $n$  then it is neologism proof for sufficiently large  $n$ .

It is clear that the fully revealing equilibrium is neologism proof, because  $\theta^1$  and  $\theta^3$  never prefer to mimic (or to be in a group with) any other type and thus have no incentive to use a neologism, which in turn means that  $\theta^2$  does not have a credible neologism as it has to involve pooling with either  $\theta^1$  or  $\theta^3$ , or both.

Let us assume (16) holds and consider when the binary communication equilibrium  $\{\theta^1\}, \{\theta^2, \theta^3\}$ , which exists for any  $n$ , is neologism proof. Clearly  $\theta^1$  does not have any incentive to use a neologism. We can also see that  $\theta^2$  does not use a neologism to separate

himself from  $\theta^3$  since

$$\left| \frac{1}{2} - \underbrace{\left( \frac{\alpha^0}{\alpha^0 + 1} \frac{7}{16} + \frac{1}{\alpha^0 + 1} \frac{7}{12} \right)}_{\bar{\mu}_2^{\text{bin}}(2)} \right| < \left| \frac{1}{2} - \underbrace{\left( \frac{\alpha^0}{\alpha^0 + 1} \frac{7}{16} + \frac{1}{\alpha^0 + 1} \frac{7}{12} - \frac{1}{\alpha^0 + n} \left( \frac{7}{12} - \frac{1}{2} \right) \right)}_{\substack{\bar{\mu}_2^{\text{bin}}(2) \\ \text{expected action when using the neologism, from (11)}}} \right|,$$

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,  $\theta^2$  is better off pooling with  $\theta^3$  than to reveal his type by a neologism.

Meanwhile,  $\theta^3$  clearly wishes to separate him from  $\theta^2$  by a neologism. However, it is not credible if  $\theta^2$  also wishes to convince the decision maker that he is  $\theta^3$ , which can be written

$$\left| \frac{1}{2} - \underbrace{\left( \frac{\alpha^0}{\alpha^0 + 1} \frac{7}{16} + \frac{1}{\alpha^0 + 1} \frac{7}{12} \right)}_{\bar{\mu}_2^{\text{bin}}(2)} \right| > \left| \frac{1}{2} - \underbrace{\left( \frac{\alpha^0}{\alpha^0 + 1} \frac{7}{16} + \frac{1}{\alpha^0 + 1} \frac{7}{12} - \frac{1}{\alpha^0 + n} \left( \frac{7}{12} - 1 \right) \right)}_{\bar{\mu}_2^{\text{bin}}(2)} \right|. \quad (21)$$

For  $\alpha^0 > 4/3$ , (21) simplifies to

$$n > \frac{-3(\alpha^0)^2 + 14\alpha^0 + 10}{3\alpha^0 - 4}. \quad (22)$$

Therefore, the neologism proofness of the binary equilibrium for a particular pair of  $\alpha^0$  and  $n$  can be checked by looking at (21) or alternatively (22), which also implies that if  $\alpha^0 > 4/3$  then *the binary equilibrium is neologism proof for large enough  $n$* . If  $\alpha^0 = 4/3$  then (21) is never satisfied, which implies there is no neologism-proof 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, which enables  $\theta^3$  to send a credible neologism.

Finally it is straightforward to see that the "babbling" equilibrium is not neologism proof. With respect to the uninformative equilibrium, a respondent with  $\theta^1$  has the incentive to separate himself because, if his message is believed, it lowers the policy towards his ideal. Meanwhile,  $\theta^2$  and  $\theta^3$  prefer the policy induced in the uninformative equilibrium, namely  $y = 7/16$ , to the policy induced by mimicking  $\theta^1$  because the policy in the uninformative equilibrium is lower than their ideal policy ( $1/2$  and  $1$ ) and mimicking  $\theta^1$  makes the policy even lower. Therefore,  $\theta^1$  has a credible neologism and hence the "babbling" equilibrium is not neologism proof.

## References

- [1] Austen-Smith, David (1993) "Interested Experts and Policy Advice: Multiple Referrals under Open Rule." *Games and Economic Behavior* 5, 3-43.
- [2] Battaglini, Marco (2002) "Multiple Referrals and Multidimensional Cheap Talk." *Econometrica* 70, 1379-1401.
- [3] ——— (2004) "Policy Advice with Imperfectly Informed Experts. *Advances in Theoretical Economics* 4, No.1 Article 1.
- [4] BBC (2011) *Scottish Election: BBC Poll Shows Scots' Top Priorities*, BBC News, 11 April 2011. Available at <http://www.bbc.co.uk/news/uk-scotland-scotland-politics-13029253>
- [5] Berinsky, Adam J. (2004) "Can We Talk? Self-Presentation and the Survey Response." *Political Psychology* 25, 643-659.
- [6] Casella, Alessandra and Gelman, Andrew (2008) "A Simple Scheme to Improve the Efficiency of Referenda." *Journal of Public Economics* 92, 2240-2261.
- [7] Crawford, Vincent P., and Sobel, Joel (1982) "Strategic Information Transmission." *Econometrica* 50, 1431-1451.
- [8] DeGroot, Morris H. (1970) *Optimal Statistical Decisions*, McGraw-Hill.
- [9] De Jong, Martijn G., Steenkamp, Jan-Benedict E. M., Fox, Jean-Paul, and Baumgartner, Hans (2008) "Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation," *Journal of Marketing Research* 45, 104-115.
- [10] Dickey, James M., Jiang, Jhy-Ming and Kadane, Joseph B. (1987) "Bayesian Methods for Censored Categorical Data." *Journal of American Statistical Association* 82, 773-781.
- [11] Farrell, Joseph (1993) "Meaning and Credibility in Cheap-Talk Games." *Games and Economic Behavior* 5, 514-531.
- [12] Feddersen, Timothy, and Pesendorfer, Wolfgang (1997) "Voting Behavior and Information Aggregation in Elections with Private Information." *Econometrica* 65, 1029-1058.
- [13] ——— and ——— (1998) "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting." *American Political Science Review* 92, 23-35.

- [14] Gerardi, Dino, and Yariv, Leeat (2008) "Information Acquisition in Committees." *Games and Economic Behavior* 62, 436-459.
- [15] Gershkov, Alex, and Szentes, Balázs (2009) "Optimal Voting Schemes with Costly Information Acquisition." *Journal of Economic Theory* 144, 36-68.
- [16] 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.
- [17] ——— and ——— (1989) "Asymmetric Information and Legislative Rules with a Heterogeneous Committee." *American Journal of Political Science* 33, 459-90.
- [18] Goeree, Jacob and Großer, Jens (2007) "Welfare Reducing Polls." *Economic Theory* 31, 51-68.
- [19] Greenleaf, Eric A. "Measuring Extreme Response Style." *Public Opinion Quarterly* 56, 328-351.
- [20] Ishida, Junichiro, and Shimizu, Takashi (2013) "Asking One Too Many? Why Leaders Need to Be Decisive" working paper.
- [21] Hu, Nan, Pavlou, Paul A., and Zhang, Jie "Overcoming the J-shaped Distribution of Product Reviews." *Communications of the ACM* 52, 144-147.
- [22] Kawamura, Kohei (2011) "A Model of Public Consultation: Why is Binary Communication So Common?." *Economic Journal* 121, 819-842.
- [23] Krishna, Vijay, and Morgan, John (2001) "A Model of Expertise." *Quarterly Journal of Economics* 116, 747-775.
- [24] Koriyama, Yukio, and Szentes, Balázs (2009) "A Resurrection of the Condorcet Jury Theorem." *Theoretical Economics* 4, 227-252.
- [25] Le Quement, Mark. (2009) "Mutual Corroboration." working paper.
- [26] Morgan, John, and Stocken, Phillip C. (2008) "Information Aggregation in Polls." *American Economic Review* 98, 864-896.
- [27] Mukhopadhyaya, Kaushik (2003) "Jury Size and the Free Rider Problem." *Journal of Law, Economics and Organization* 19, 24-44.
- [28] Ottaviani, Marco, and Sørensen, Peter N. (2001) "Information Aggregation in Debate: Who Should Speak First?" *Journal of Public Economics* 81, 393–421.

- [29] Persico, Nicola (2004) "Committee Design with Endogenous Information." *Review of Economic Studies* 71, 165-191.
- [30] Rothschild, Michael (1974) "Searching for the Lowest Price when the Distribution of Prices is Unknown." *Journal of Political Economy* 82, 689-711.
- [31] Shapiro, Robert Y. (2011) "Public Opinion and American Democracy." *Public Opinion Quarterly* 75, 982-1017.
- [32] Talmain, Gabriel (1992) "Search from an Unknown Distribution: An Explicit Solution." *Journal of Economic Theory* 57, 141-157.
- [33] Wolinsky, Asher (2002) "Eliciting Information from Multiple Experts." *Games and Economic Behavior* 41, 141-160.