Identification of misreported beliefs

Elias Tsakas*
Maastricht University

December 2021

Abstract

It is well-known that subjective beliefs cannot be identified with traditional choice data unless we impose the strong assumption that preferences are state-independent. This is seen as one of the biggest pitfalls of incentivized belief elicitation. The two common approaches are either to exogenously assume that preferences are state-independent, or to use intractable elicitation mechanisms that require an awful lot of hard-to-get non-traditional choice data. In this paper we use a third approach, introducing a novel methodology that retains the simplicity of standard elicitation mechanisms without imposing the awkward state-independence assumption. The cost is that instead of insisting on full identification of beliefs, we seek identification of misreporting. That is, we elicit beliefs with a standard simple elicitation mechanism, and then by means of a single additional observation we can tell whether the reported beliefs deviate from the actual beliefs, and if so, in which direction they do.

Keywords: Belief elicitation; misreporting; state-dependent preferences.
JEL codes: C91, C93, D80, D81, D82, D83.

1. Introduction

Being able to obtain unbiased estimates of people's beliefs is of outmost importance for explaining and predicting behavior, and designing policy interventions (Manski, 2004). However, given the inherent latency of beliefs, obtaining such estimates has to rely heavily on (self-)reporting. Unfortunately, in practice, people often misreport their beliefs, even when they are incentivized to tell the truth. Thus, an important question is to identify whether their reports indeed deviate from their actual beliefs, and if so, in which direction.

As an illustration of the problem, take the usual finding: Democrats and Republicans systematically disagree — usually by a large margin — on the probability they reportedly assign to some politically charged event, e.g., the winner of the next elections (Bullock et al., 2015). The literature on politically motivated beliefs would say that the two groups actually have different beliefs. However, there is an alternative explanation: actual beliefs are not that divergent, and differences are amplified by misreporting (in the direction of their preferred parties respectively). Making this distinction is crucial, e.g., when deciding on whether to regulate (mis-)information. This is because, actual belief polarization — as opposed to mere exaggeration of reported beliefs — has the potential to trigger extreme political reactions.

Another common observation is that most people report that they are themselves more skilled than their average peer, e.g., when asked about their driving skills (Svenson, 1981). Once again, the

*Department of Economics (MPE), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; Homepage: www.elias-tsakas.com; E-mail: e.tsakas@maastrichtuniversity.nl

literature on motivated beliefs would say that this is consistent with the actual beliefs being biased in favor of the preferred state. But once again, one could argue that this is because people misreport their beliefs about their own perceived ability. And again, the distinction can have important consequences, e.g., for insurance purposes. If people actually overestimate their abilities, they may end up taking more risks when they drive, as opposed to situations where they just exaggerate their self-reported abilities.

But why would people misreport even when they are incentivized to tell the truth? A number of psychological factors have been recently proposed in the literature, such as self-image (Ewers and Zimmermann, 2015), preference to appear truthful to their audience (Thaler, 2021), deliberate attempt to express attitudes, known as "cheerleading" (Bullock et al., 2015; Hannon and de Ridder, 2021). The common thread among these explanations is that people have state-dependent preferences, i.e., they have some sort of stakes in the realization of the underlying event.

Among theorists this is not particularly surprising: it is well known for years that, whenever preferences are state-dependent, beliefs cannot be identified using only data on traditional choices among acts (Fishburn, 1973; Karni et al., 1983; Drèze, 1987). In particular, even if we somehow observed the complete preference relation (over acts), it would still be the case that for every belief there would exist a (state-dependent) utility function such that the resulting expected utility function would represent these preferences. So, as colossal as Savage's (1954) subjective expected utility theory is, it will only help us to identify beliefs if preferences are exogenously assumed to be state-independent.[1] And of course, the same problem is inherited by almost all belief elicitation mechanisms, viz., proper scoring rules (Brier, 1950; Good, 1952; Savage, 1971), binarized scoring rules (Hossain and Okui, 2013), matching probabilities (Ducharme and Donnell, 1973; Kadane and Winkler, 1988; Baillon et al., 2018), clock auctions (Karni, 2009; Tsakas, 2019), promisory notes (De Finetti, 1974; Kadane and Winkler, 1988). This is because all these mechanisms essentially boil down to choosing from menus of acts, i.e., they rely on traditional choice data.

As a response to this conundrum, two approaches have been taken in the literature, which we will call the "practically oriented" and the "theoretically sound" approach.

According to the practically oriented approach, it is still assumed, in the spirit of Savage (1954) and Anscombe and Aumann (1963), that preferences are state-independent. The idea is pretty pragmatic: existing elicitation mechanisms are simple and easy to implement. This is what makes them appealing after all. Thus, we are keen to maintain this simplicity even if it comes at the price of imposing an exogenous structural assumption (viz., state-independence).

On the other hand, the theoretically sound approach dispenses the awkward state-independence assumption. However, this means that we need to go well beyond traditional choice data, in ways that make belief elicitation practically intractable. In this sense, it is not surprising that this literature is quite thin (Karni, 1999; Jaffray and Karni, 1999) and has not been adopted by empirical researchers.[2]

It is not hard to guess that we find neither of the two approaches very satisfactory. So, we propose a third alternative, which maintains the simplicity of the existing elicitation tasks, without at the same time imposing the awkward state-independence assumption. The price that we will pay for reconciling the two is that we no longer insist on full identification of beliefs, but rather we only seek identification of misreporting. In other words, we will be able to tell if the reported beliefs — that we have already elicited using our standard simple techniques — deviate from the actual beliefs, and if so in which direction. For instance, in our earlier example, we will not be able to pin down the exact

[1] The same holds true for the famous subsequent attempt of Anscombe and Aumann (1963).

[2] There is also large related literature within axiomatic decision theory. However, the different conditions that lead to identification of beliefs (and utilities) are perhaps even more demanding. Of course, from the point of view of this literature, this is not a concern, as the overall aim is to show that identification is in principle feasible and therefore the notion of subjective beliefs is well-defined. We further elaborate on the specific contributions in the literature section.

beliefs of each Democrat and each Republican, but we will be able to tell which ones exaggerate when they report their forecasts. Thus, we will be able to test whether the difference in their beliefs has been amplified by the fact that they both have stakes in the event they are forecasting.

Our method is on a high level inspired by the moral hazard literature (Drèze, 1987; Drèze and Rustichini, 1999; Baccelli, 2021), in that we exploit the presence of some action which is known to affect the agent's belief in a certain direction.[3] We call such an action *influential*. Of course, the crucial difference is that in the moral hazard literature the influential action is controlled by the agent herself, whereas in our case it is controlled by the analyst (Remark 2). For instance, in the context of our earlier example, an influential action would be a donation to the campaign of the Democratic candidate, or the initiation of a negative rumor for the Republican candidate, in some swing state. In fact, any action that helps one of the two candidate would do the trick, even if the help is marginal (see Examples 2-3). Although we cannot quantify the effect of each of these actions on an agent's beliefs, we can safely assume that there will be an increase in the probability of the Democratic candidate winning.

Then, our method proceeds as follows. First, we elicit beliefs using a proper binarized scoring rule, i.e., an incentive-compatible scoring rule that pays in probabilities to win a fixed prize.[4] Subsequently, we ask the agent to choose between two fifty-fifty lotteries, whose outcomes are combinations of whether the prize is paid and whether the influential action is taken. In the context of the previous example, the two lotteries can be labeled as a "risky option" and a "hedging option". The risky option is a coin toss that either will pay both the prize to the agent and the donation to the campaign, or will not pay any of the two. On the other hand, the hedging option is a coin toss that either will only pay the prize to the agent, or will only pay the donation to the campaign. Then, our main result shows that the choice between the two lotteries identifies misreporting in the preceding belief elicitation task (Theorem 1). In particular, if the risky option (resp., the hedging option) is chosen, the reported probability of the Democratic candidate winning is greater (resp. smaller) than the actual belief. If the subject is indifferent between the two options, then the reported belief is the same as the actual belief, i.e., there is no misreporting.

The bottomline is that our approach allows us to keep using the state-of-the-art elicitation methodology, and only adds on top a simple task which identifies whether the agent has misreported beliefs.[5] Thus, simply put, we make a significant step in solving a long-standing problem (viz., belief elicitation under state-dependent preferences) at a very small cost (viz., adding a single question to the current methodology). Of course, our solution is partial, but in many applied settings — where a qualitative analysis is used — full identification is anyway not needed. Besides, if our method concludes that there is no misreporting, full identification is achieved.

The only papers that introduce mechanisms for eliciting beliefs under state-dependent preferences are Karni (1999) and Jaffray and Karni (1999), with the latter proposing two different mechanisms. In particular, Karni (1999) and the first mechanism of Jaffray and Karni (1999) rely on the assumption that state utilities are bounded, and they approximate the actual beliefs in the limit as monetary incentives grow arbitrarily large.[6] This is a rather uncomfortable convention, as the elicitation task will rely on a very large dataset. Moreover, we will either need to incur an extremely high cost, or to use hypothetical data. These problems are recognized by the authors of the two aforementioned papers, who point out that in those early days of the literature there was no other option (e.g., Karni,

[3]The term "moral hazard" should not be confused with the one used in information economics.

[4]Later in the paper, we show that our analysis holds verbatim for any incentive-compatible binarized elicitation mechanism, e.g., binarized matching probabilities, or clock auctions (Section 4.2). The common feature of all these mechanisms is that they do not need to impose any assumption on the subject's risk preferences, which is what makes them very appealing. We also discuss the extension to non-binarized mechanisms, such as arbitrary proper scoring rules (Section 4.1).

[5]In Section 4.3, we explain that distortions due to hedging opportunities are not really a concern.

[6]For an extensive discussion on the boundedness of the utility function, see Wakker (1993).

1999, p.485). The second mechanism in Jaffray and Karni (1999) assumes that state-dependence enters the picture in terms of unobserved state-dependent payments. So, first, it proceeds to elicit these payments, and once these are known, it goes on to elicit beliefs using standard techniques. Of course, this is a rather restrictive setting: in most interesting applications, preferences over states are intrinsic. Besides, in order to elicit the state-dependent payments is quite demanding in terms of the amount of data that is needed.

As we have already mentioned, there is also a large literature within axiomatic decision theory. The various attempts to identify beliefs (without exogenously assuming state-independence) differ in terms of the additional choice domain — beyond traditional choice data — that they employ, and of course on the corresponding axioms they impose. For instance, Fishburn (1973) allows for comparison between acts conditional on different events. Karni et al. (1983) and Karni and Schmeidler (2016) introduce hypothetical preferences over acts, conditional on exogenously given probabilities over the states. Schervish et al. (1990) allow the agent to compare lotteries at different states. Drèze (1987) and Drèze and Rustichini (1999) allow for the agent to be able to influence the state realization, in different ways depending on the act she faces. Lu (2019) introduces stochastic choices under different information structures. For a more complete account of this literature, we refer to the reviews of Drèze and Rustichini (2004), Karni (2008), and more recently Baccelli (2017). Of course, the aim of this whole literature is anyway to establish that beliefs are well-founded and that they can be in principle identified, rather than to suggest concrete methods for actually eliciting said beliefs. In this sense, it is not surprising that using these representation results to actually identify beliefs would be quite a challenge.

Finally, our work is methodologically similar to the one of Offerman et al. (2009), who first elicit beliefs using standard proper scoring rules, and then design a subsequent test that identifies misreporting due to violations of risk-neutrality and/or presence of probability weighting.

The paper is structured as follows: Section 2 presents the relevant background concepts. In Section 3 we introduce our mechanism and present our results. In Section 4 we discuss extensions and implementation.

2. Background

2.1. State-dependent subjective expected utility

Take a binary state space $\Theta = \{\theta_0, \theta_1\}$. Probability measures over Θ are identified by the probability they attach to θ_1. Let $\mathcal{L}_X$ be the set of lotteries over a set $X \subseteq \mathbb{R}$ of monetary payoffs. Moreover, let $\mathcal{F} = \mathcal{L}_X^{\Theta}$ denote the set of acts, with typical element f. Consider a (female) agent who maximizes subjective expected utility (abbrev., SEU). That is, there exist a state-dependent (strictly increasing) Bernoulli utility function $u = (u_0, u_1)$ and a belief $\mu \in (0,1)$, such that her preferences over $\mathcal{F}$ are represented by the function

$$\mathbb{E}_\mu(u(f)) := (1-\mu)u_0\big(f(\theta_0)\big) + \mu u_1\big(f(\theta_1)\big), \tag{1}$$

where $u_0(f(\theta_0)) := \langle f(\theta_0), u_0\rangle$ and $u_1(f(\theta_1)) := \langle f(\theta_1), u_1\rangle$ are the (vNM) expected utilities that the lotteries $f(\theta_0)$ and $f(\theta_1)$ yield at states θ_0 and θ_1 respectively. We will say that the SEU representation is state-independent, if $u_0 = u_1$.

As it is well-known this representation is not unique. Indeed, for an arbitrary belief $\tilde{\mu} \in (0,1)$, the pair $(\tilde{u}, \tilde{\mu})$ is also a subjective expected utility (SEU) representation of the same preferences, if we set

$$\tilde{u}_0 = \frac{1-\mu}{1-\tilde{\mu}} u_0 \quad \text{and} \quad \tilde{u}_1 = \frac{\mu}{\tilde{\mu}} u_1. \tag{2}$$

This is because $\mathbb{E}_{\tilde{\mu}}(\tilde{u}(f)) = \mathbb{E}_{\mu}(u(f))$ for every act $f \in \mathcal{F}$. As a result, beliefs cannot be identified with traditional choice data (i.e., preferences over $\mathcal{F}$) alone. Note that identification of beliefs is impossible even if there exists a state-independent SEU representation.[7] In order to deal with this fundamental identification problem, different solutions have been proposed in the literature, relying on collecting additional data, well beyond choices over acts.

One crucial point we should stress is that throughout the paper, we will assume that there exists an actual belief and an actual utility function. Such a pair can be interpreted either as a primitive — which is actually how we prefer to view it — or as the parameters that one would obtain by using one of the aforementioned identification results. One way or another, we will say that preferences are state-independent whenever a state-independent SEU representation exists, and moreover it coincides with the *actual* SEU representation. The first part of the previous statement (i.e., existence of a state-independent representation) can be tested with traditional choice data, but the second part (i.e., the state-independent representation being the actual one) needs additional data in order to be tested.

2.2. Proper scoring rules

Scoring rules are mechanisms that aim at incentivizing the agent to report her (actual) beliefs truthfully, by rewarding her based on her reported belief and the realized state. Formally, a scoring rule is a function

$$\pi : [0,1] \to \mathcal{F}$$

that takes as an input the reported belief $r \in [0,1]$, and returns as an output the act $\pi_r \in \mathcal{F}$ that the agent receives in return. Formally speaking, a scoring rule is the menu of acts, $\{\pi_r \mid 0 \leq r \leq 1\}$. In this sense, the agent's reported belief is a single point of traditional choice data.

A scoring rule π is called binarized whenever it pays in lotteries over two fixed monetary payoffs (Hossain and Okui, 2013), i.e., for every report r and every state θ, the lottery $\pi_r(\theta)$ is distributed over a good payoff $\overline{x}$ and a bad payoff $\underline{x}$. We will refer to the good payoff as the prize, and to the probability of winning the prize as the winning probability.

A scoring rule is proper if reporting truthfully (uniquely) maximizes the total expected payoff (given the actual beliefs) over the set of all possible reports $r \in [0,1]$. Obviously, for a binarized scoring rule, maximizing the total expected payoff is equivalent to maximizing the total winning probability.

The appeal of properness is that it claims to identify the agent's beliefs using a single observation of traditional choice data. However — given the earlier identification problem — it is not surprising that this cannot be done, unless we impose additional assumptions on the utility functions (Kadane and Winkler, 1988; Karni and Safra, 1995). To see why this is the case, suppose that some $p \in (0,1)$ has been reported in response to the scoring rule, and observe that there exist infinitely many SEU representations, some with beliefs $\mu < p$ and some others with beliefs $\mu > p$. In principle, we do not know which of these representations corresponds is the actual one. So, in order to identify the actual beliefs we would need to somehow exogenously restrict the set of SEU representations. The way this is typically done is by imposing exogenous assumptions on the utility function. For instance, when a binarized scoring rule is used, it is implicitly assumed that preferences are state-independent. On the other hand, when an arbitrary — non-binarized — scoring rule is used, even stronger assumptions are needed, i.e., both state-independence and risk-neutrality are implicitly assumed. The bottom line is that state-independence is always needed if we want to maintain incentive-compatibility. And this has been recognized as perhaps the biggest pitfall of incentivized belief elicitation.

[7] Recall that a state-independent SEU representation exists whenever the monotonicity axiom of Anscombe and Aumann (1963) is satisfied.

3. Identifying deviations from actual beliefs

From our previous discussion it follows that, if we want to fully identify the agent's beliefs, we will eventually face a fundamental tradeoff. Namely, we will need either to have a rich dataset (going well beyond traditional choice data), or to exogenously assume state-independent preferences. This tradeoff is well-known among theorists, but is often overlooked in practice, where we simply use proper scoring rules without further discussion on the possibility of preferences being state-dependent.

Here we will take a different approach. We will maintain both the principle of a "minimal dataset", and we will dispense with the assumption of state-independent preferences. However, in order to be able to accommodate both of these requirements simultaneously, we will relax full identification. In particular, instead of aiming to pin down the agent's actual beliefs, we simply want to learn if the agent has misreported or not. And if she has, we also want to know which direction she has deviated. Notably, we will do all this at the expense of only one additional observation (besides the belief report).

Inspired on a high level by the moral hazard literature (Drèze, 1987; Drèze and Rustichini, 1999; Baccelli, 2021), suppose that we can influence the state realization. In particular, assume that there exists some action $\hat{a}$ available to ourselves (viz., the experimenters), which is commonly known to affect the likelihood of θ_1 in a certain direction. Throughout the paper, we will refer to $\hat{a}$ as the *influential action*, and without loss of generality, we will assume that increases the probability of θ_1 to some $\hat{\mu} > \mu$. Notably, we remain agnostic on how much the belief will increase in response to the influential action: all we know is that it will increase. Not picking the influential action $\hat{a}$ means that we stick to the default action a which would leave the agent's beliefs unaffected to μ. Here are a couple of examples of influential actions:

Example 1. We are interested in the beliefs of a Democrat about the Democratic candidate winning the upcoming elections. One influential action would be to donate an amount to the Democratic campaign. Another influential action would be to commit some additional votes in a swing state to this candidate (assuming of course that this is a credible commitment). A third influential action would be to start a rumor on social media that the Democratic candidate will increase minimum wages and will decrease taxes. Note that this last action would not involve deception: the influential action is not the realization of the rumor (i.e., the increase of wages or the decrease of taxes) but rather the rumor itself.[8] In either case, the agent's subjective probability of the Democratic candidate winning will increase. ◁

Example 2. We are interested in an investor's beliefs about a company going bankrupt before the end of the current year. One influential action would be for us to invest money in this company. Another influential action would be to start a rumor that the company is about to file for new patent.[9] In both cases, it is reasonable to assume that the investor's subjective belief of bankruptcy will go down. ◁

Example 3. We are interested in the beliefs of a young economist about her paper being published in a top journal. One influential action would be to put a good word with a friendly editor. Another influential action would be to commit that a prominent economist will carefully read the manuscript and provide comments before the paper is submitted. In both cases, the subjective probability the author assigns to the paper being accepted will go up. ◁

Remark 1. It is really important to choose an influential action which does not affect the agent directly, besides the effect that it has on the state space. For instance in Example 3, we should not

[8] As a disclaimer, we are not recommending experiments that spread fake news. We only use it as an example to illustrate how an influential action functions.

[9] Here the same comment (regarding deception) applies as in the previous example.

try to influence the editor if the author of the paper has ethical issues with lobbying. This is because in such case, the influential action would distort not only the agent's beliefs, but also her utilities. ◁

Notice that in our case, it is us (viz., the experimenters) who control the influential action, as opposed to the moral hazard literature where the action is controlled by the agent. As a result, we can construct lotteries over the product space $X \times \{a, \hat{a}\}$. These are not usual lotteries that pay only in monetary payoffs. Instead, an outcome of such a lottery would be a pair of a monetary payoff (which affects the agent directly) and an action (which affects the agent indirectly via the uncertainty on Θ).

For the two monetary outcomes, $\underline{x}$ and $\overline{x}$, define the lotteries:

$$\begin{aligned} A &:= \left(\frac{1}{2} \times (\underline{x}, a), \frac{1}{2} \times (\overline{x}, \hat{a})\right), \\ B &:= \left(\frac{1}{2} \times (\overline{x}, a), \frac{1}{2} \times (\underline{x}, \hat{a})\right). \end{aligned} \tag{3}$$

Intuitively, in the context of Example 1, suppose that the good payoff is $\overline{x} = \$10\text{k}$, while the influential action is a donation of $\hat{a} = \$10\text{k}$. Then, A can be seen as a "risky option" for the agent, in the sense that either \$20k will be paid out in total (\$10k to herself and \$10k to the campaign), or no money at all will be paid out. On the other hand, B can be seen as a "hedging option" for the agent, in the sense that \$10k will be paid out anyway, either to the agent herself or to campaign.

Remark 2. By having a choice between A and B, the agent cannot influence the state realization. This is because regardless which of the two lotteries is chosen, the influential action and the default action will both occur with probability one half. This is a major difference with the moral hazard literature, which relies on the agent being able to affect the state. ◁

Then, the following results use the agent's revealed preferences over the pair of lotteries to identify misreporting, viz., A is chosen (resp., B is chosen) if the reported belief is above (resp., below) the actual beliefs.

Theorem 1. *Let μ be the agent's actual beliefs, and p be her reported beliefs in response to a proper binarized scoring rule. Then, we have $p > \mu$ (resp., $p < \mu$), if and only if, $A \succ B$ (resp., $A \prec B$).*

Proof. First, denote by $\pi_r^k := \pi_r(\theta_k)(\overline{x})$ the winning probability (at state θ_k) when the report r is submitted. By properness of π, we have

$$\begin{aligned} (1-r)\pi_r^0 + r\pi_r^1 &> (1-r)\pi_p^0 + r\pi_p^1, \\ (1-p)\pi_r^0 + p\pi_r^1 &< (1-p)\pi_p^0 + p\pi_p^1. \end{aligned}$$

For any $r < p$, we have $\pi_r^0 < \pi_p^0$ and $\pi_r^1 > \pi_p^1$, and therefore

$$\frac{r}{1-r} < \frac{\pi_p^0 - \pi_r^0}{\pi_r^1 - \pi_p^1} < \frac{p}{1-p}, \tag{4}$$

whereas for any $r > p$, we have $\pi_r^0 > \pi_p^0$ and $\pi_r^1 < \pi_p^1$, and therefore we obtain

$$\frac{p}{1-p} < \frac{\pi_p^0 - \pi_r^0}{\pi_r^1 - \pi_p^1} < \frac{r}{1-r}. \tag{5}$$

Taking side limits of in (4) and (5) as r approaches p from the below and above respectively, yields

$$\lim_{r\uparrow p}\frac{\pi_p^0-\pi_r^0}{\pi_r^1-\pi_p^1}=\lim_{r\downarrow p}\frac{\pi_p^0-\pi_r^0}{\pi_r^1-\pi_p^1}=\frac{p}{1-p}. \tag{6}$$

Now, given the actual belief μ, the expected utility from the report r is equal to

$$\mathbb{E}_\mu\big(u(\pi_r)\big)=(1-\mu)\Big(\pi_r^0u_0(\overline{x})+\big(1-\pi_r^0\big)u_0(\underline{x})\Big)+\mu\Big(\pi_r^1u_1(\overline{x})+\big(1-\pi_r^1\big)u_1(\underline{x})\Big). \tag{7}$$

By p being actually reported, it follows that $\mathbb{E}_\mu\big(u(\pi_p)\big)\geq\mathbb{E}_\mu\big(u(\pi_r)\big)$ for every $r\neq p$. This means that whenever it is the case that $r<p$, we have

$$\frac{\mu}{1-\mu}\cdot\frac{u_1(\overline{x})-u_1(\underline{x})}{u_0(\overline{x})-u_0(\underline{x})}\geq\frac{\pi_p^0-\pi_r^0}{\pi_r^1-\pi_p^1}, \tag{8}$$

while whenever it is the case that $r>p$, we obtain

$$\frac{\mu}{1-\mu}\cdot\frac{u_1(\overline{x})-u_1(\underline{x})}{u_0(\overline{x})-u_0(\underline{x})}\leq\frac{\pi_p^0-\pi_r^0}{\pi_r^1-\pi_p^1}. \tag{9}$$

Hence, if we take the side limits in (8) and (9) as r approaches p from below and above respectively, and we use Equation (6), it will follow that

$$\frac{p}{1-p}=\frac{\mu}{1-\mu}\cdot\frac{u_1(\overline{x})-u_1(\underline{x})}{u_0(\overline{x})-u_0(\underline{x})}. \tag{10}$$

Obviously, this directly implies the equivalence

$$p\geq\mu\;\Leftrightarrow\;\frac{u_1(\overline{x})-u_1(\underline{x})}{u_0(\overline{x})-u_0(\underline{x})}\geq 1, \tag{11}$$

with the first inequality being strict, if and only if, the second one is strict.

Now, let $\hat{\mu}>\mu$ be the unobserved probability that the agent attaches to θ_1 if $\hat{a}$ is chosen. Then, the following equivalences hold:

$$\begin{aligned}
A\succeq B\;&\Leftrightarrow\;\frac{1}{2}\mathbb{E}_\mu\big(u(\underline{x})\big)+\frac{1}{2}\mathbb{E}_{\hat{\mu}}\big(u(\overline{x})\big)\geq\frac{1}{2}\mathbb{E}_\mu\big(u(\overline{x})\big)+\frac{1}{2}\mathbb{E}_{\hat{\mu}}\big(u(\underline{x})\big)\\
&\Leftrightarrow\;\mathbb{E}_{\hat{\mu}}\big((u(\overline{x})-u(\underline{x})\big)\geq\mathbb{E}_\mu\big((u(\overline{x})-u(\underline{x})\big)\\
&\Leftrightarrow\;(\hat{\mu}-\mu)\big(u_1(\overline{x})-u_1(\underline{x})\big)\geq(\hat{\mu}-\mu)\big(u_0(\overline{x})-u_0(\underline{x})\big)\\
&\Leftrightarrow\;\frac{u_1(\overline{x})-u_1(\underline{x})}{u_0(\overline{x})-u_0(\underline{x})}\geq 1,
\end{aligned}$$

with the last inequality being strict, if and only if the preference relation is strict. Combining this last equivalence with (11) completes the proof. □

Remark 3. In case the influential action is known to decrease — rather than increase — the subjective probability of θ_1, the previous result still stands verbatim with the preference ordering reversed, i.e., $p>\mu$, if and only if, $A\prec B$. ◁

Note that the previous results crucially rely on us being able to set the probabilities in each of the two lotteries to exactly fifty-fifty. Let us explain why fifty-fifty probabilities are so crucial. First, note that the agent will over-report her belief of θ_1, if and only if, the utility function is (locally) supermodular, i.e., formally speaking, the difference $u_1 - u_0$ is increasing as we move from $\underline{x}$ to $\overline{x}$ (see (11)).[10] Then, we go on to show that this supermodularity condition is characterized by the preferences over these exact fifty-fifty lotteries.[11] Let us illustrate why this is the case. Suppose that the payments are $\underline{x} = 0$ and $\overline{x} = 1$. Since the influential action leads to an increased probability $\hat{\mu} > \mu$, it will be the case that $u_1 - u_0$ is increasing, if and only if, $\mathbb{E}_{\hat{\mu}}(u(\cdot)) - \mathbb{E}_{\mu}(u(\cdot))$ is increasing. Then, take the two straight lines, one that connects the graph of $\mathbb{E}_{\hat{\mu}}(u(\cdot))$ evaluated at 1 to the graph

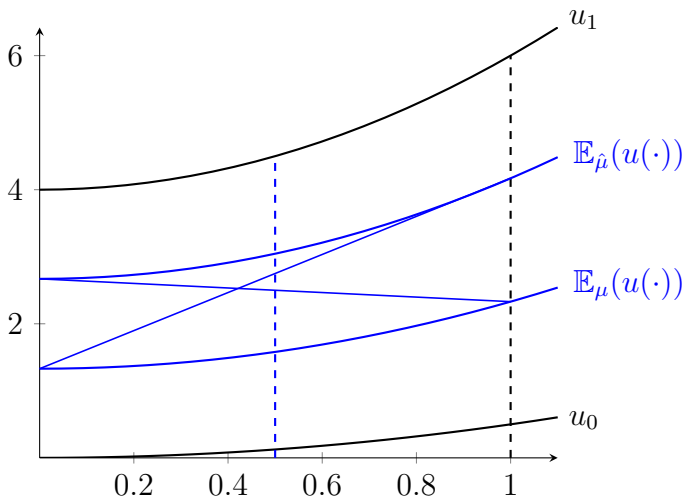

Figure 1: Supermodularity is characterized by preferences over $\{A, B\}$.

of $\mathbb{E}_{\mu}(u(\cdot))$ evaluated at 0, and one that connects the graph of $\mathbb{E}_{\mu}(u(\cdot))$ evaluated at 1 to the graph of $\mathbb{E}_{\hat{\mu}}(u(\cdot))$ evaluated at 0. Finally, observe that the difference $\mathbb{E}_{\hat{\mu}}(u(\cdot)) - \mathbb{E}_{\mu}(u(\cdot))$ in increasing, if and only if, these two lines intersect to left of $1/2$. But then again, the two lines intersect to the left of $1/2$, if and only if, A is preferred to B.

An alternative approach would have been to induce fifty-fifty probabilities via an information structure that yields two signals, each occurring with probability a half. In particular, suppose that we can construct an experiment that yields either signal s_0 (which is known to increase the probability of θ_0) or signal s_1 (which is known to increase the probability θ_1). Then, we would be asking the subject whether she prefers to be paid the good payoff $\overline{x}$ when s_0 is realized and the bad payoff $\underline{x}$ when s_1 is realized, or vice versa (similarly to Lu, 2019). However, this alternative mechanism would rely on two very strong assumptions. First, we would need to make sure that we can design such an experiment. However, this would be practically impossible unless we knew the prior μ, which of course we do not know. Second, if we were hypothetically able to design such an experiment, we would need to know that the agent updates in a Bayesian manner. This we do not know either. So overall, as theoretically appealing as this alternative mechanism may look, the implementation would be rather difficult.

[10] The same condition is obtained by Kadane and Winkler (1988) and Jaffray and Karni (1999) for matching probabilities. This implies that our method works verbatim if we replace binarized scoring rules with matching probabilities. In fact, the same is true for other elicitation methods (see Section 4.2).

[11] Interestingly, a similar condition is used by Francetich (2013) in his characterization result of supermodular vNM EU functions.

4. Discussion

4.1. Non-binarized scoring rules

The reason binarized scoring rules are appealing is because they do not require any assumption — besides state-independence — in order to guarantee truth-telling. This is in contrast to arbitrary proper scoring rules (e.g., the commonly-used quadratic scoring rule) which need to assume risk-neutrality — on top of state-independence — in order to retain incentive-compatibility. Of course, we should note that there is a debate on the tradeoff between incentive-compatibility and not needing reduce compound lotteries, e.g., see Selten et al. (1999), Harrison et al. (2013), Harrison et al. (2014), Harrison et al. (2015), just to mention a few. Although we personally find the overall evidence to favor binarized scoring rules, it is not our aim to participate in this discussion. Instead we ask the following question: if we assume risk-neutrality at both states, can we identify misreporting due to state-independence? The answer is affirmative: Using Theorem 1 for any two payments $\underline{x}$ and $\overline{x}$ identifies at which of the two states the marginal utility is greater, which in turn reveals misreporting. In this sense, our method is not restricted to the binarized case.

4.2. Beyond scoring rules

As we have already mentioned, our methodology holds verbatim if we replace binarized scoring rules with any other binarized elicitation task, such as matching probabilities or clock auctions. The reason is that the belief elicitation task is essentially independent of our additional task that identifies misreporting. For instance, similarly to our analysis of scoring rules, Kadane and Winkler (1988) and Jaffray and Karni (1999) show that matching probabilities will induce misreporting in favor of θ_1, if and only if, the difference $u_1 - u_0$ is increasing. Then, the choice between our lotteries A and B characterizes the monotonicity of $u_1 - u_0$. Hence, our method will tell us whether the reported belief deviates from the actual one, and if so, in which direction.

4.3. Hedging

A well-known concern regarding incentivized belief elicitation is the possibility of hedging (Blanco et al., 2010). In general terms this means that the agent can choose a optimal strategy for the grand decision problem which includes both the elicitation task and some other task, which does not induce truth-telling in the elicitation task. The most common manifestation of the problem is due to non-risk-neutral risk preferences. In our case, this is not really a problem, as long as we pay randomly for one of the two tasks, i.e., either the binarized scoring rule or the choice from $\{A, B\}$. Crucially, the chosen lottery in the second task will be realized only if the second task has been drawn to be compensated. This is done so that beliefs in the first task are not distorted in anticipation of the possibility that the influential action will be drawn in the second task.

References

Anscombe, F.J. & Aumann, R.J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics* 34, 199–205.

Baccelli, J. (2017). Do bets reveal beliefs? A unified perspective on state-dependent utility issues. *Synthese* 194, 3393–3419.

——— (2021). Moral hazard, the Savage framework, and state-dependent utility. *Erkenntnis* 86, 367–387.

Baillon, A., Huang, Z., Selim, A. & Wakker, P. (2018). Measuring ambiguity attitudes for all (natural) events. *Econometrica* 86, 1839–1858.

Blanco, M., Engelmann, D., Koch, A. & Normann, H.T. (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics* 13, 412–438.

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78, 1–3.

Bullock, J., Gerber, A., Hill, S. & Huber, G. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science* 10, 519–578.

De Finetti, B. (1974). *Theory of probability.* Vol. 1, Wiley.

Drèze, J. (1987). Decision theory with moral hazard and state-dependent preferences. *Essays on Economic Decisions under Uncertainty* 23–89.

Drèze, J. & Rustichini, A. (1999). Moral hazard and conditional preferences. *Journal of Mathematical Economics* 31, 159–181.

——— (2004). State-dependent utility and decision theory. *Handbook of Utility Theory*, Ch. 8, 839–892.

Ducharme, W. & Donnell, M. (1973). Intrasubject comparison of four response modes for subjective probability assessment. *Organizational Behavior and Human Performance* 10, 108–117.

Ewers, M. & Zimmermann, F. (2015). Image and misreporting. *Journal of teh European Economic Association* 13, 363–380.

Fishburn, P. (1973). A mixture-set axiomatization of conditional subjective expected utility. *Econometrica* 41, 1–25.

Francetich, A. (2013). Notes on supermodularity and increasing differences in expected utility. *Economics Letters* 121, 206–209.

Good, I.J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B*, 14, 107–114.

Hannon, M. & de Ridder, J. (2021). The point of political belief. *Routledge Handbook of Political Epistemology* (forthcoming).

Harrison, G., Martìnez-Correa, J. & Swarthout, J.T. (2013). Inducing risk neutral preferences with binary lotteries: a reconsideration. *Journal of Economic Behavior and Organization* 94, 145–159.

——— (2014). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior and Organization* 101, 128–140.

Harrison, G., Martìnez-Correa, J., Swarthout, J.T. & Ulm, E. (2015). Eliciting subjective probability distributions with binary lotteries. *Economics Letters* 127, 68–71.

Hossain, T. & Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies* 80, 984–1001.

Jaffray, J.Y. & Karni, E. (1999). Elicitation of subjective probabilities when the initial endowment is unobservable. *Journal of Risk and Uncertainty* 8, 5–20.

Kadane, J. & Winkler, R. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association* 83, 357–363.

Karni, E. (1992). Subjective probabilities and utilities with event-dependent preferences. *Journal of Risk and Uncertainty* 5, 107–125.

——— (1993). A definition of subjective probabilities with state-dependent preferences. *Econometrica* 61, 187–198.

——— (1999). Elicitation of subjective probabilities when preferences are state-dependent. *International Economic Review* 40, 479–486.

——— (2008). State-dependent utility. *Handbook of Rational and Social Choice*, 223–238.

——— (2009). A mechanism for eliciting probabilities. *Econometrica* 77, 603–606.

Karni, E. & Safra, Z. (1995). The impossibility of experimental elicitation of subjective probabilities. *Theory and Decision* 38, 313–320.

Karni, E. & Schmeidler, D. (2008). An expected utility theory for state-dependent preferences. *Theory and Decision* 81, 467–478.

Karni, E., Schmeidler, D. & Vind, K. (1983). On state-dependent preferences and subjective probabilities. *Econometrica* 51, 1021–1031.

Lu, J. (2019). Bayesian identification: a theory for state-dependent utilities. *American Economic Review* 109, 3192–3228.

Manski, C.F. (2004). Measuring expectations. *Econometrica* 72, 1329–1376.

Offerman, T., Sonnemans, J., van de Kuilen, G. & Wakker, P. (2009). A truth serum for non-Bayesians: correcting proper scoring rules for risk attitudes. *Review of Economic Studies* 76, 1461–1489.

Savage, L. (1954). *The foundations of statistics.* Wiley, NY: Dover Publications.

——— (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* 66, 783–801.

Schervish, M., Seidenfeld, T. & Kadane, J. (1990). State-dependent utilities. *Journal of the American Statistical Association* 85, 840–847.

Selten, R., Sadrieh, A. & Abbink, K. (1999). Money does not induce risk neutral behavior, but binary lotteries do even worse. *Theory and Decision* 46, 211–249.

Svenson, O. (1981). Are we all less risky and more skillfull than our fellow drivers?. *Acta Psychologica* 47, 143–148.

Thaler, M. (2021). The supply of motivated beliefs. *Working Paper.*

Tsakas, E. (2019). Obvious belief elicitation. *Games and Economic Behavior* 118, 374–381.

Wakker, P. (1993). Unbounded utility for Savage's "foundations of statistics", and other models. *Mathematics of Operations Research* 18, 446–485.