

Belief identification with state-dependent utilities*

ELIAS TSAKAS[†]

Maastricht University

March, 2022

Abstract

It is well known that individual beliefs cannot be identified using traditional choice data, unless we impose the practically restrictive and conceptually awkward assumption that utilities are state-independent. In this paper, we propose a novel methodology that solves this long-standing identification problem in a simple way. Our method relies on the concept of influential actions. These are actions that are controlled by the analyst and lead the agent to change her beliefs. Notably, the analyst does not need to have any idea on how the agent’s beliefs will change in response to an influential action. Then, instead of eliciting directly the agent’s beliefs about the state space, we elicit her subjective probabilities about the influential action having been undertaken conditional on each state realization. The latter can be easily done with existing elicitation tools. It turns out that this is enough to uniquely identify her beliefs about the state space irrespective of her utility function, thus solving the identification problem. We discuss that this method can be used in most applications of interest. As an example, we show how it can provide a new useful tool for identifying motivated reasoning on an individual level.

KEYWORDS: Belief identification; state-dependent utility; influential action; elicitation mechanisms; motivated beliefs.

JEL CODES: C91, C93, D80, D81, D82, D83.

1. Introduction

1.1. Problem statement

Identifying subjective beliefs is a central problem in economics that dates back to the early seminal contributions of Ramsey (1931), De Finetti (1937) and Savage (1954). While this literature originally focused mostly on providing foundations for the notion of subjective probability, more recently economists have also recognized the practical importance of the question (Manski, 2004). This renewed interest comes mainly from the fact that beliefs are nowadays widely used as an explanatory variable for behavior. Therefore, being able to measure beliefs accurately is of outmost importance for applied research.

The traditional methodology relies on identifying beliefs through observed betting behavior. Specifically, an agent’s choices among acts (i.e., state-contingent lotteries) are supposed to reveal

*An earlier version of this paper circulated under the title “Identification of misreported beliefs”. I am greatly indebted to Jean Baccelli, Aurelien Baillon, Edi Karni, Jay Lu, Andy Mackenzie and Peter Wakker for their instrumental comments. I would also like to thank seminar and conference audiences for useful input.

[†]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

her beliefs over the state space (Savage, 1954; Anscombe and Aumann, 1963; Wakker, 1989). For instance, if we observe the agent’s willingness to accept different betting odds on the Democratic candidate winning the next presidential elections, we will be able to pin down her exact beliefs about the election outcome.

However, there is a major caveat, known as the *identification problem*, which undermines the significance of this methodology (e.g., Drèze, 1961, 1987; Fishburn, 1973; Karni et al., 1983):

The agent’s beliefs can be identified through traditional choices over acts, only if we exogenously assume her utility function to be state-independent.

That is intuitively, we must assume that the agent values material payoff in exactly the same way at all states. Obviously, this is a very restrictive assumption: in most applications where behavior crucially depends on beliefs, utility is state-dependent, e.g., insurance problems (Drèze and Rustichini, 2004), legal judgments (Andreoni, 1991; Feddersen and Pesendorfer, 1998; Tsakas, 2017), medical decisions (Pauker and Kassirer, 1975, 1980), etc. Furthermore, this assumption implies that the agent is indifferent between any two states, thus ruling out at the outset phenomena like motivated reasoning (Kunda, 1990; Benabou, 2015), which are widely studied in behavioral economics and psychology.

The identification problem was already noticed during the very early days of decision theory: as Leonard Savage admits in his well-known letter correspondence with Bob Aumann, “the problem is serious, but I am willing to live with it until something better comes along” (Aumann and Savage, 1971). What is hidden behind these simple words of Savage is a fundamental tradeoff. Namely, in order to identify an agent’s beliefs we must either accept the conceptually awkward assumption of state-independent utilities, or we must go well beyond traditional choice data. The former approach is usually adopted by experimental economists, who—in the name of simplicity—disregard the problem. And although at the outset this may sound as a simplistic approach, it is actually quite pragmatic. Indeed, even though decision-theorists have proposed various extensions of the choice domain to circumvent the identification problem (e.g., Drèze, 1961, 1987; Fishburn, 1973; Karni et al., 1983; Karni, 1992; Lu, 2019), none of them is unanimously accepted as the standard one. This suggests that the identification problem is both very hard and still open.

1.2. Our contribution

Actually, the main reason why none of the aforementioned theoretically-sound attempts to overcome the identification problem has been broadly adopted by applied researchers is that they all require—to different extent—complex and rich data. And this is exactly where our contribution lies. Namely, we introduce a novel method, which allows us to identify beliefs in an easy and tractable way, without needing to impose any assumption on the agent’s utility function. In this sense, we provide a new theoretical solution to the long-standing identification problem, which has the crucial advantage that it can be used directly off the shelf for applied purposes.

The key ingredient of our method is what we call an *influential action*. This is an action that can be undertaken by the analyst, and leads the agent to somehow revise her belief. Note that the analyst does not need to know anything about how the agent’s beliefs will change in response to this action, which is what makes the use of influential actions appealing. For instance, recall our earlier example where the analyst is interested in the agent’s belief about the outcome of the next presidential election. By the way, this is a case where the identification problem will likely arise, e.g., if the agent is a Democrat herself, she will most likely have a preference for the Democratic candidate to win. One influential action would be to make a significant donation to the Democratic campaign. In this case, the agent’s belief about the Democrats winning would most certainly change.¹ If on the

¹Many more examples are presented in detail in Section 3. In fact, as we will be explaining, in most applications

other hand, no donation was made, the agent would not revise her beliefs, i.e., she would maintain the same beliefs that we have been trying to identify all along.

So, here is how we leverage the possibility of using the influential action in order to pin down these beliefs. First of all, given that it is up to the analyst whether the donation is made or not, he can choose himself the probability with which money is donated. This probability is publicly announced, and will be henceforth thought as the agent’s prior probability of the influential action being taken, e.g., suppose that the analyst informs the agent that the probability of donation is 20%, and the probability of no donation is 80%. Importantly, the realization of this lottery is not revealed to the agent, i.e., the agent does not know yet whether the donation has been actually made or not. Then, before the agent learns who won the election, she reports two posterior probabilities: her subjective probability that the donation has been made given the Democrat winning,

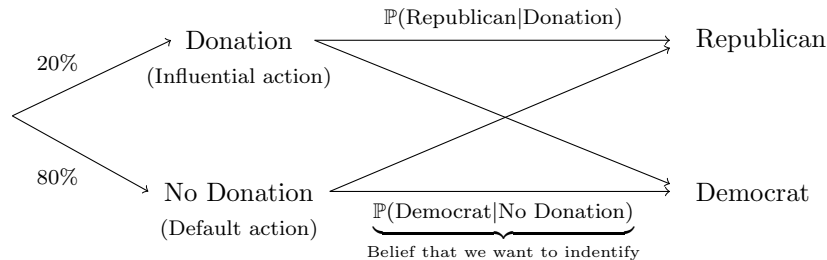
$$\mathbb{P}(\text{Donation}|\text{Democrat}) = 30\%,$$

and her subjective probability that the donation has been made given the Republican winning,

$$\mathbb{P}(\text{Donation}|\text{Republican}) = 10\%.$$

For the reports of her posterior probabilities, the agent is rewarded with some incentive-compatible elicitation task from the rich belief elicitation literature.²

Knowing the prior and posterior probabilities, we can directly retrieve the likelihoods, i.e., the agent’s beliefs about the election winner given that no donation has been made.



But this is exactly the belief that we wanted to identify in the first place. More specifically, by simply applying Bayes rule, we obtain the following equations:

$$\underbrace{\mathbb{P}(\text{No Donation}|\text{Republican})}_{0.90} = \frac{0.80\mathbb{P}(\text{Republican}|\text{No Donation})}{0.80\mathbb{P}(\text{Republican}|\text{No Donation}) + 0.20\mathbb{P}(\text{Republican}|\text{Donation})},$$

$$\underbrace{\mathbb{P}(\text{No Donation}|\text{Democrat})}_{0.70} = \frac{0.80\mathbb{P}(\text{Democrat}|\text{No Donation})}{0.80\mathbb{P}(\text{Democrat}|\text{No Donation}) + 0.20\mathbb{P}(\text{Democrat}|\text{Donation})}.$$

Finally, solving this system yields the agent’s belief without donation,

$$\mathbb{P}(\text{Democrat}|\text{No Donation}) = 7/16,$$

which is the belief that we wanted to identify all along. In this sense, we have managed to identify the agent’s belief with only two data points, i.e., the reported posterior probabilities about the influential action. This is the essence of our main result (Theorem 1).³

there will typically exist a multitude of influential actions, many of which will be much cheaper and much easier to implement than the donation. In this sense, our method is going to be very general.

²The choice of the elicitation task is orthogonal to the rest of our method: see our discussion in Section 5.2.

³Notably, our result identifies, not only the agent’s beliefs given “No Donation, but also her beliefs given “Donation”. Thus, we can also pin down how the agent internalizes information due to the influential action being taken. So, as an important side result, this provides us with a new methodology for identifying motivated reasoning: see Section 6 for a detailed discussion.

What makes our method circumvent the identification problem is that, when we elicit the agent’s posterior probabilities about the influential action (viz., about the donation), we condition on the state (viz., on the Democrats having won, and then again on the Republicans having won). So, utilities become by construction state-independent. Moreover, the influential action has been chosen in a way such that we can safely assume that it will not affect the utilities directly. For instance, it is natural to assume that the agent does not care per se about a donation being made. Of course, she cares indirectly, as the donation improves the chances of the Democrats winning the election. But this only affects the likelihoods, rather than the utility function. As a result, we can use standard identification techniques—which assume state-independent utilities—from the extensive belief elicitation literature in order to pin down these posterior probabilities.

The theoretical contribution of the previous approach is straightforward: we manage to solve the long-standing identification problem in a simple way without any assumption on the utility function. The only question that remains then is whether this method is easily applicable. Practically, the two main potential obstacles stem from our assumption that the agent updates in a Bayesian manner and from our confidence that a suitable influential action can be found. We address both of them in Section 5.

Regarding the updating issue, we extend our main identification result to allow for non-Bayesian updating (Theorem 2). This would require additional out-of-sample data to calibrate updating parameters, which we can then use to identify the agent’s beliefs. Methodologically, this approach is similar to the one in [Offerman et al. \(2009\)](#), in the sense that we keep using our original simple mechanism and then we debias the elicited beliefs ex post.

On the other hand, regarding the issue of finding an appropriate influential action, one should first notice how broad our definition of influential actions is. In particular, the fact that we can remain agnostic on how a certain influential action affects the agent’s beliefs, means that we can choose from a very broad range of actions. Actually, most of them are classified in one of two broad categories. First, there are those that affect the realization of the state, like for instance the donation in our earlier example. These influential actions can only be used in cases where the true state has not been realized yet. The alternative is to present the agent with some evidence, while letting her know that it is true with certain probability and fake otherwise. In this case, the influential action is the true evidence. For instance, again in the previous example, suppose that the analyst has conducted an election poll, according to which the Democratic candidate will receive 60% of the votes. If the agent takes into account this poll, it is safe to assume that her belief about the winner would change. However, what the agent sees is the true poll with probability 20%, and a randomly-generated poll with probability 80%. And once again, this is known to the agent. Note that this second type of influential actions can also be used in cases where the state has been already realized, but remains unknown to the agent. As we also previously mentioned in Footnote 1, additional examples (of both types) are presented in Section 3.

1.3. Literature review

There is a lot of existing work dealing with the identification problem. This literature is roughly split in two streams, one that focuses on providing tools which can be potentially used in practice for belief elicitation purposes, and another one which is seen as part of decision theory and focuses mostly on the conditions under which belief identification can be achieved theoretically. For a more complete account of this literature, we refer to the reviews of [Drèze and Rustichini \(2004\)](#), [Karni \(2008, 2016\)](#) and [Baccelli \(2017\)](#).

Starting with the first stream, 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.⁴ 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, 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, eliciting the state-dependent payments is quite demanding in terms of the amount of data that is needed.

Turning to the second stream, the various attempts within axiomatic decision theory differ in terms of the non-traditional choice domain they consider. Within this stream, there are three main methodological approaches, all of which rely on the agent’s beliefs being somehow revised at some instance.

The first such method, which was mainly followed in the early days, was based on exogenously manipulating context and a fortiori the agent’s beliefs. 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. Karni (1992, 1993) allows the analyst to observe preferences conditional on different events. The main difference of this work in comparison to ours is that they all assume the analyst to influence the agent’s beliefs in a way that he knows and controls. Moreover, typically, the way the agent’s beliefs are influenced is either not specified or hypothetical. These two features make the implementation of these methods cumbersome, to say the least.

The second method relies on the idea that the agent herself can affect the state realization, and is called the *moral hazard approach*. In fact, it originally appeared in the early sixties (Drèze, 1961), before resurfacing again a couple of decades later (Drèze, 1987; Drèze and Rustichini, 1999) and recently receiving attention again (Baccelli, 2021). This literature relates to our work on a high level. Of course the difference is that according to this approach the analyst needs to rely on the agent influencing the state realization, whereas in our case the analyst has full control of the situation. The second major difference is that the moral hazard methodology applies only to settings where the state has not been realized yet, while our method does not pose any such restriction (see our earlier discussion on the different categories of influential actions).

The third and final method is more recent, and relies on the agent updating her beliefs using information that the analyst provides (Lu, 2019). This is admittedly a very promising method: similarly to ours, it can be potentially applied both in cases where the state has been already realized, as well as in cases where it has not. It does not impose restrictive assumptions at the outset. The only potential drawback that we see, is that it requires stochastic choices (under different information structures), implying that one would need a rather large number of observations. Nevertheless, we still see it as a very promising alternative methodology.

Somewhere in between the second and the third method, one can place a sequence of papers that rely on the idea that the agent can influence the state realization and choose an act, conditional on different signals (Karni, 2011a,b, 2016). The common element across these papers of Karni papers, Lu’s aforementioned paper and our work is that all can be thought to rely on usual choice data over an expanded state space. Nevertheless, the way this general idea is implemented is very different.

⁴For an extensive discussion on the boundedness of the utility function, see Wakker (1993).

2. The identification problem

In the formal setting of [Anscombe and Aumann \(1963\)](#), we begin with a binary state space $\Theta = \{\theta_0, \theta_1\}$ and a set of outcomes X .⁵ Without loss of generality and unless mentioned otherwise, we assume X to be finite. Probability measures over Θ are identified by the probability they attach to θ_1 . Let \mathcal{L}_X be the set of lotteries over X , and \mathcal{L}_X^Θ denote the set of acts. That is, act $f \in \mathcal{L}_X^\Theta$ maps state $\theta_k \in \Theta$ to a lottery $f_k \in \mathcal{L}_X$ for each $k \in \{0, 1\}$.

A female agent has a subjective belief $\mu \in (0, 1)$ and a state-dependent utility function $u = (u_0, u_1)$ where $u_k : X \rightarrow \mathbb{R}$ is the state-utility function.⁶ Both beliefs and utilities are considered to be primitives: The belief is a measure of how likely she deems each state, while the utility is a description of her tastes. We say that utility is state-independent if $u_0 = u_1$. Otherwise, it is state-dependent. Utility can be state-dependent, for instance, due to state-dependent monetary stakes (e.g., a side bet) or because of intrinsic preferences over the state space (e.g., preference for winner of elections). Notice that utility being state-independent is a stricter requirement than the vNM preferences over \mathcal{L}_X being state-independent, e.g., if u_1 is a positive affine transformation of u_0 , the vNM preferences will be state-independent, whereas the utility function will not.

The agent evaluates acts using *subjective expected utility (SEU)*.⁷ That is, each act $f \in \mathcal{L}_X^\Theta$ is assigned the corresponding value

$$\mathbb{E}_\mu(u(f)) := (1 - \mu) \sum_{x \in X} f_0(x)u_0(x) + \mu \sum_{x \in X} f_1(x)u_1(x), \quad (1)$$

and the agent's preferences over acts are given by the usual equivalence, viz.,

$$\text{act } f \text{ is preferred to act } g \text{ if and only if } \mathbb{E}_\mu(u(f)) \geq \mathbb{E}_\mu(u(g)).$$

Thus, we say that preferences over acts are represented by the pair (u, μ) .⁸ Unfortunately, this is only one of the infinitely many SEU representations. For instance, for any belief $\tilde{\mu} \in (0, 1)$ and rescaled utility function

$$\tilde{u}_0 = \frac{1 - \mu}{1 - \tilde{\mu}} u_0 \quad \text{and} \quad \tilde{u}_1 = \frac{\mu}{\tilde{\mu}} u_1, \quad (2)$$

the pair $(\tilde{u}, \tilde{\mu})$ will constitute an alternative SEU representation. This is because we will have $\mathbb{E}_{\tilde{\mu}}(\tilde{u}(f)) = \mathbb{E}_\mu(u(f))$ for every act $f \in \mathcal{L}_X^\Theta$. So, even if we hypothetically manage to get access to the complete preference relation over acts, we will not be able to tell if her belief is μ or $\tilde{\mu}$. That is, formally speaking, beliefs cannot be identified from traditional choice data (i.e., from observed choices over acts). This is known as the *identification problem* of SEU.

Let us elaborate a bit. First of all, note that traditional choice data allow us to test whether the agent evaluates acts using a SEU function. This follows from the fact that preferences over acts admit a SEU representation if and only if they satisfy the standard vNM axioms (viz., weak order, continuity, independence). But at the same time, even if the vNM axioms are indeed satisfied, the preferences will admit way too many SEU representations. In fact, there will exist so many SEU representations that every possible belief will appear in one of these representations. So, the only way to uniquely identify beliefs from traditional choice data is by exogenously restricting the set of SEU representations in terms of introducing assumptions on the utility function.

⁵Our analysis can also be done in the framework of [Savage \(1954\)](#).

⁶Throughout the paper, for exposition simplicity, we restrict focus on full-support beliefs. This assumption is without loss of generality: see [Section 7.1](#) for an extension to boundary beliefs.

⁷Note that we use SEU as a general term that includes both the state-independent and the state-dependent version.

⁸Using the terminology that we have already established, (u, μ) is a state-independent SEU representation if $u_0 = u_1$, and it is a state-dependent SEU representation otherwise.

The typical such assumption postulates that utilities are state-independent: if we assume $u_0 = u_1$ and $\tilde{u}_0 = \tilde{u}_1$ in (2), we will directly obtain $\mu = \tilde{\mu}$. Unfortunately, this assumption cannot be corroborated with traditional choice data. Namely, we can tell if one of the infinitely many SEU representations is state-independent, but we cannot rule out alternative state-dependent representations.⁹ For instance, in (2), even if there is a state-independent SEU representation (u, μ) , there will also exist additional state-dependent SEU representations $(\tilde{u}, \tilde{\mu})$ that prescribe different beliefs. So, the identification problem persists, unless we assume it away.

The identification problem explains why traditional elicitation mechanisms are completely uninformative to the analyst, unless one imposes strong assumptions on the utility function, like for instance state-independence. First, note that a traditional *belief elicitation mechanism* is a direct mechanism that incentivizes the agent by rewarding accurate reports and punishing inaccurate ones. Formally, it can be written as a function

$$\pi : [0, 1] \rightarrow \mathcal{L}_X^\Theta \quad (3)$$

that takes as input the agent's reported belief $r \in [0, 1]$, and returns as output the act $\pi(r) \in \mathcal{L}_X^\Theta$ that the agent will receive in return, i.e., the agent's reward is a lottery that depends on her report and the realized state.¹⁰ An elicitation mechanism is *incentive-compatible* within a class U of utility functions, if it guarantees truth-telling for every utility function within this class, i.e., for all $u \in U$ it is the case that

$$\mathbb{E}_\mu(u(\pi(\mu))) > \mathbb{E}_\mu(u(\pi(r))), \quad (4)$$

for every $\mu \in [0, 1]$ and every $r \neq \mu$. As it turns out, there are many incentive-compatible mechanisms if we restrict the set U so that it contains only the state-independent utility functions. Examples of such mechanisms include all proper binarized scoring rules, the Karni mechanisms, and mechanisms that use the method of matching probabilities.¹¹ In fact, throughout the paper, whenever we say that an elicitation mechanism is incentive-compatible (without specifying U), we will mean that it is incentive-compatible with respect to the set of state-independent utility functions.

However, the flip side is that every elicitation mechanism is essentially a menu of acts. So, without a strong exogenous assumption on the utility function (like state-independence), the agent's choice from this menu will not suffice to identify her beliefs. In other words, formally speaking, there is no incentive-compatible mechanism for the universal set U that contains all utility functions.

3. Influential actions

From our previous discussion it follows that, if we want to identify the agent's beliefs, we will eventually face a fundamental tradeoff. Namely, we will need either to go well beyond traditional choice data, or to impose strong exogenous assumptions on the utility function.

This tradeoff is well-known already since the early days of decision theory (Drèze, 1961; Aumann and Savage, 1971). Unsurprisingly, theorists have taken the first approach and have proposed various

⁹Checking whether there is a state-independent SEU representation can be done by testing, on top of the usual vNM axioms, the monotonicity axiom in the framework of Anscombe and Aumann (1963) or axioms P3 and P4 in the framework of Savage (1954). These tests will not guarantee that all SEU representations are state-independent.

¹⁰Existing mechanisms in the literature differ in how the incentives are framed. For instance, scoring rules (Brier, 1950; Good, 1952; Savage, 1971) present these incentives explicitly by providing the exact function π to the agent. On the other hand, the Karni mechanisms (Karni, 2009; Tsakas, 2019) and the method of matching probabilities (Ducharme and Donnell, 1973; Baillon et al., 2018) associate each r with a compound act, which can in turn be reduced into a simple act π_r .

¹¹If we further restrict U to contain only linear state-independent utility functions, the usual quadratic scoring rule becomes incentive-compatible too.

types of non-traditional choice domains, but the fact that there is no broad consensus suggests that none of these approaches is unanimously appealing. One of the main drawbacks from which all these attempts seem to suffer—to a different extent—is that they all require quite complex and rich data, and therefore belief elicitation in practice turns into a rather difficult endeavor. In this sense, it is not that surprising that although the decision-theoretic literature on this topic is reasonably broad, the corresponding belief elicitation literature is very thin (Karni, 1999; Jaffray and Karni, 1999). Perhaps, this also explains why experimental economists have taken the more pragmatic approach of ignoring the identification problem and simply using traditional elicitation mechanisms, i.e., in the name of simplicity, they implicitly assume utilities to be state-independent, so that truth-telling is guaranteed.

In this paper, we introduce a novel methodology which will allow us to circumvent the identification problem in a simple way without needing to make such exogenous assumptions on the utility function. In particular, suppose that the analyst can actively influence the agent’s beliefs, by undertaking some action a_ν which leads to a change of the agent’s subjective belief from μ to some $\nu \in (0, 1)$. Throughout the paper, we will refer to a_ν as the *influential action*.¹² Notably, we do not place any restriction on how the influential action affects the agent’s beliefs:

(A₁) The analyst remains agnostic on how the agent’s belief will change in response to the influential action. All we know is that it will be different from the original one, i.e., $\nu \neq \mu$.

Not picking the influential action means that the analyst sticks to the default action a_μ , which can be thought as “the analyst doing nothing”. This would leave the agent’s beliefs unaffected to μ . Importantly, the probability of picking the influential action is a choice variable of the analyst, and all that we will eventually require is that the agent knows this probability.

There are two general types of influential actions. The first one is inspired on a high level by the moral hazard literature (Drèze, 1961, 1987; Drèze and Rustichini, 1999; Baccelli, 2021), in the sense that the action influences the realization of Θ . In this sense, influential actions of this type can be used in settings where the state has not been realized yet. Of course, the crucial difference is that in our framework it is the analyst who influences the state space, whereas in the moral hazard literature it is the agent herself. This difference turns out to be crucial. Here are a couple of examples of this type of influential actions:

Example 1. The analyst is 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). In both these cases, it is safe to assume that the agent’s subjective probability of the Democratic candidate winning will change, as both actions help the Democrat win. ◀

Example 2. The analyst wants to identify an investor’s beliefs about a company going bankrupt before the end of the current year. One influential action would be to invest money in this company. Then, it is reasonable to assume that the investor’s subjective belief of bankruptcy will go down. ◀

Example 3. The analyst is interested in the beliefs of a young economist about her paper being published in a top journal. One influential action would be to assign a friendly editor to handle the paper. This would change the author’s subjective probability of her paper being accepted. ◀

Example 4. A car manufacturer wants to know the probability that their own employee assigns to the event that the new model will be approved by the Environmental Protection Agency. One influential action could be that the manufacturer falsifies some of the test results on the emissions the car is producing. This would increase the probability that the model is approved. ◀

¹²Influential actions were first introduced in the predecessor of this paper (Tsakas, 2020).

Example 5. A train company wants to know the beliefs of free-riders that they will get inspected. One influential action is to hire 10 more inspectors, which will make it more likely that passengers are checked. Another influential action is to shift most inspectors to work during peak hour, which would also influence the beliefs of the passengers somehow, although it is unclear in which way this will happen as we do not necessarily know when they travel during the day. \triangleleft

As a final side remark on this first approach, it is possible that an influential action triggers a subsequent action by the agent herself —like in the moral hazard literature— that also affects her beliefs. For instance, in Example 3, shifting more inspectors to peak hour may lead the passenger to start travelling more off peak hours, which also affects her subjective probability of being caught. The overall effect on her belief is unclear, but this is not a problem for our methodology. All we care about is that our influential action causes an eventual overall belief change. Whether this is attributed only to the influential action itself or also to subsequent reactions to this influential action is not our concern.

The second type of influential actions is useful in settings where a state has already been realized, but the agent remains uncertain about it. Obviously, in these cases we cannot undertake an action that influences the realization of Θ , like for instance influential actions of the first type do. So, what we do instead is to present the agent with some evidence which is true with some probability and it is fabricated otherwise. Crucially, the probability of the evidence being true is chosen by the analyst. If the evidence is true, the agent will change her beliefs to some ν . On the other hand, if it is fabricated, she will maintain her current belief μ . In this sense, true evidence can be seen as the default action, and fabricated evidence as the influential one. This type of influential actions relates on a high level to the stream of literature that identifies beliefs using different information structures (Lu, 2019). Let us present some examples of this type of influential actions:

Example 1 (continued). Recall our earlier example, where the analyst wants to know the beliefs of a Democrat about the Democratic candidate winning the elections. Suppose that —instead of trying to influence the realization of the state space— the agent is presented with the results of a poll according to which 60% of the respondents will vote Democrat. The agent is told that these results come from a true poll with probability 1/2 and they are completely fabricated with probability 1/2. In this case, the influential action is that the poll results are true, in which case the agent’s beliefs change. On the other hand, the default action is that the poll results are fabricated, in which case her beliefs remain the same. \triangleleft

Example 6. We want to identify a juror’s belief about the defendant being actually guilty. This is clearly a setting where the juror has state-dependent utilities, as she would rather convict a guilty defendant and acquit an innocent one. Suppose that we tell the juror that there is some camera feed that shows the defendant near the scene of the murder. The influential action is that the evidence is indeed true, in which case the juror’s belief about the defendant being guilty will change. The default action is that the evidence is fabricated, in which case her beliefs will remain the same. \triangleleft

Example 7. We want to identify the subjective probability that a driver’s assigns to herself being more skilled than the median driver. By the way, this is one of the very early settings where motivated beliefs were studied (Svenson, 1981). Assume that we provide some new statistics on last year’s accident rates. If this data is true, the influential action has been taken, and the driver will most likely use it to update her perception of the quality of the median driver, and a fortiori it will influence her belief on her relative performance compared to the median driver. If, on the other hand, this data is fabricated, the default action has been realized, and her beliefs will not be influenced. \triangleleft

The analysis that we will carry throughout the paper will be the same irrespective of which type of influential actions we use. Thus, we will not make this distinction explicit henceforth.

There is only one assumption that we need to impose for our identification method to work:

(A₂) The influential action does not affect the agent directly, besides the indirect effect that it has on her beliefs.

Formally, this means that the utility function may vary across Θ , but not across A . For instance, in Example 3, the author of the paper only cares about whether the paper is published or not, irrespective of who the editor is. Likewise, in Example 6, the juror is not affected by the feed being true or not, other than the influence this will have on her beliefs. The idea is that, without this assumption, the influential action would distort not only the agent’s beliefs, but also her utility function. Fortunately, in most applications there will exist multiple influential actions, and therefore we can safely assume that the analyst can find one that does not affect the agent directly.

From a technical point of view, introducing an influential action can be seen as enlarging the state space, by adding a new dimension of uncertainty, i.e., the agent will now face the extended state space $\Theta \times A$. Her belief over $\Theta \times A$ is latent, but its marginal over A is known. Then, the analyst can now collect non-traditional data on choices whose outcome depends on both Θ and A . As we will see in the upcoming sections, such data suffice for identification of the agent’s conditional beliefs about Θ given the action a_μ , which is exactly the belief μ that the analyst wanted to elicit in the first place. This will be achieved by leveraging the fact that the analyst controls the probability of the influential action, and a fortiori knows the agent’s marginal belief over A . Although the general framework seems complex at first glance, our elicitation mechanisms will be framed as simple choice problems which are easy to describe and implement.

4. Full identification of beliefs

We will now introduce our main mechanism, which fully identifies the agent’s beliefs. This will be done using an indirect approach that elicits beliefs about A . The mechanism works as follows.

Mechanism 1. Let μ and ν denote the agent’s latent beliefs given the default action a_μ and the influential action a_ν respectively.

STEP 1: The analyst randomly picks an action from $A = \{a_\mu, a_\nu\}$. The agent knows that the default action a_μ is drawn with probability $p \in (0, 1)$, but she does not know which action is actually drawn.

STEP 2: Before learning the true state, the agent’s conditional beliefs about A are elicited given each of the two states in Θ . In particular:

STEP 2.1: She reports the probability p_0 of a_μ conditional on θ_0 .

STEP 2.2: She reports the probability p_1 of a_μ conditional on θ_1 .

For each θ_k , the agent is rewarded with an incentive-compatible mechanism $\pi_k : [0, 1] \rightarrow \mathcal{L}_X^A$.

STEP 3: The realized state in Θ and the realized action in A are revealed, and the agent is paid only for the accuracy of her report given the realized state.

Before moving forward, let us emphasize once more that the elicitation mechanisms in Step 2 are defined over a different state space compared to the one we introduced in Section 2.¹³ In particular,

¹³Our mechanism bears some similarity to the one that Thaler (2020) uses in order to identify motivated reasoning: he also eventually elicits the agent’s subjective probability of a news source being truthful in order to test whether some piece of information about Θ is internalized differentially by people of different political affiliations. Still, his mechanism is fundamentally different from ours, not only in the terms of the underlying question, but also in that he needs to directly elicit the (median) belief about Θ , which based on our entire discussion might suffer from the identification problem. Furthermore, his mechanism applies only to influential actions of the second type.

since the aim of this elicitation task is to identify the agent’s beliefs about A , the reported belief p_k is about A . Moreover, the payment induced by π_k depends on the realization of A , rather than the realization of Θ which is anyway fixed by the fact that we have conditioned on θ_k , i.e., formally, π_k takes as input the agent’s (conditional) report and the realized action in A , and pays back a lottery over X (in case the state is realized). Examples of incentive compatible such mechanisms would include proper binarized scoring rules, Karni mechanisms, and mechanisms based on the method of matching probabilities. We further elaborate on the issue elicitation in Section 5.2.

Let us illustrate this mechanism in the context of Example 1. The agent (viz., the Democrat whose beliefs we want to elicit) knows that there is ex ante 10% probability that a sizeable amount will be donated to the campaign of the Democratic candidate. Then, she is asked to submit two reports: her subjective probability that the donation has been made in case the Democrats win the election, and her subjective probability that the donation has been made in case the Republicans win. She will be paid based on the accuracy of the relevant report once the actual winner is known.

Intuitively, Mechanism 1 can be interpreted as an auxiliary information structure, where A is the state space, Θ is the signal space, and the beliefs μ and ν are the likelihoods of θ_1 conditional on a_μ and a_ν respectively (Figure 1). Then, upon observing θ_k the agent updates her subjective belief of a_μ in a Bayesian manner to the posterior β_k .¹⁴

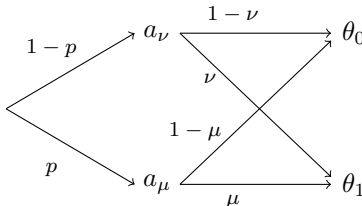


Figure 1: Mechanism 1 as an auxiliary information structure.

Our identification strategy will rely on two key observations. First, the prior beliefs over A are known by construction. Second, and perhaps most crucially, the posterior beliefs are truthfully reported, i.e., $p_k = \beta_k$ for both $k \in \{0, 1\}$. This is because, given the state θ_k , the agent’s utility function will be u_k , irrespective of which action in A has been taken, i.e., she is essentially facing an incentive-compatible elicitation task in a setting with state-independent utilities. Actually, it is exactly the fact that utility is state-independent with respect to A which guarantees that incentive-compatible elicitation mechanisms exist in the first place (see our discussion at the end of Section 2). Finally, knowing the prior and the posterior beliefs allows us to identify the likelihoods, i.e., the beliefs μ and ν .

Theorem 1 (FULL IDENTIFICATION OF BELIEFS). *Using Mechanism 1, if the agent updates in a Bayesian manner, her beliefs are identified by*

$$\mu = \frac{p_1(p - p_0)}{p(p_1 - p_0)} \quad \text{and} \quad \nu = \frac{(1 - p_1)(p - p_0)}{(1 - p)(p_1 - p_0)}. \quad (5)$$

Interestingly, the previous result allows us to identify not only the belief μ , in which we were interested in the first place, but also the belief ν that the agent would hold conditional on the influential action. This means that we can use the result to elicit how the agent interprets information. For instance, we can identify how people with different political preferences internalize information with a politically-charged content (e.g., a given election poll). This is a question that has attracted a lot of attention within the extensive literature on motivated reasoning, and our result provides an

¹⁴Later on, when we discuss implementation issues, we relax the assumption on Bayesian updating (Section 5.1).

appealingly simple answer to it (see Section 6). This is an important bonus that we obtain on top of our primary goal, which was to identify the belief μ .

Summarizing, from a theoretical point of view, the importance of the contribution is clear: we provide a novel solution to a long-standing fundamental problem that dates all the way back to Savage (1954). And in fact, we do so with minimal input in terms of underlying assumptions. Furthermore, in terms of the amount of data, only two choices are needed to identify two beliefs, which is the absolutely minimum one could have hoped. Finally, with regards to the nature of the data, it all relies upon the existence of an influential action. But then again, based on how permissive we are in what constitutes an influential action, we should be fairly certain that in most settings influential actions exist. In fact, all the examples in Section 3 support this optimism. So, concluding, unlike most contributions within decision theory, our mechanism is simple enough to become the basis for an elicitation methodology that can be used directly off the shelf.

5. Belief elicitation in practice

As we have already discussed, part of the appeal of our method comes from its simplicity, which in turn makes it a strong candidate for being used for practical purposes. Even so, research has shown that implementation of elicitation tasks often presents obstacles. Here we would like to point out potential such issues, and explain how they can be (easily) addressed and mitigated in our context.

5.1. Belief updating

The main issue that we have to confront is that we rely on the assumption that the agent updates beliefs in a Bayesian manner. Although from a theoretical standpoint this is not difficult to defend, from an empirical point of view we would have to seriously address the possibility of updating biases if we were to actually use our mechanism for eliciting beliefs in an experimental setting.

The general framework within which updating biases are typically studied in the literature was introduced by Grether (1980). This model embeds Bayesian updating as a special case, and it is flexible enough to capture all known updating biases (Benjamin, 2019). Formally speaking, it introduces two free parameters $c > 0$ and $d > 0$, which distort the true likelihoods and the prior beliefs respectively, i.e., likelihoods are raised to c and priors are raised to d . Subsequently, beliefs are updated in the usual way, i.e., by taking the relative (distorted) likelihood of each state, weighted by the corresponding (distorted) priors. In our context, posterior beliefs would then be given by

$$\beta_0 = \frac{p^d(1-\mu)^c}{p^d(1-\mu)^c + (1-p)^d(1-\nu)^c} \quad \text{and} \quad \beta_1 = \frac{p^d\mu^c}{p^d\mu^c + (1-p)^d\nu^c}. \quad (6)$$

In general, c measures the inference bias, i.e., to what extent the signal is taken into account: in this sense it is called the *inference parameter*. Parameter d measures the extent to which the prior is used in the formation of the posteriors, and it is therefore called *base-rate parameter*. Obviously, whenever $c = d = 1$, we are back to the Bayesian benchmark. Moreover, note that whenever the prior is uniformly distributed, the base-rate parameter drops out, and posterior beliefs are only potentially distorted by the inference parameter. Thus, it becomes significantly easier to identify c (in the absence of d), which is why many updating experiments start with uniform prior beliefs. In our case, this is a very convenient observation, as the prior p is controlled by the experimenter, and therefore can be set to $1/2$.

Our approach to deal with the possibility of non-Bayesian updating relies on calibrating the updating parameters using out of sample data. In particular, we begin by calibrating c , as well as

d in case we have set p different than one half.¹⁵ Subsequently, we identify the agent’s beliefs using the same mechanism, while taking the calibrated parameters into account.

Theorem 2. *Using Mechanism 1, if the agent’s updating parameters are $c > 0$ and $d > 0$, the agent’s beliefs are identified by*

$$\mu = \frac{\alpha_1(1 - \alpha_0)}{\alpha_1 - \alpha_0} \quad \text{and} \quad \nu = \frac{1 - \alpha_0}{\alpha_1 - \alpha_0}, \quad (7)$$

where $\alpha_k := \left(\frac{p_k}{1-p_k}\right)^{1/c} \left(\frac{1-p}{p}\right)^{d/c}$.

The idea is methodologically similar to the one used by Offerman et al. (2009), who in a different context choose to stick to a simple mechanism, being aware that it may yield biased measurements. Then, using out of sample data they calibrate parameters that allow them to correct these biases.¹⁶

Recall that we still assume that both μ and ν are full-support, implying that both $p_0, p_1 \in (0, 1)$. Hence, both α_0 and α_1 are well-defined. As we have already mentioned this assumption is without loss of generality. In fact, in Section 7.1, we generalize this formula to also allow for boundary reports.

Notably, given the fact that Theorem 2 provides us with closed-form solutions for the beliefs, we can use the confidence intervals around the estimated parameters c and d to obtain confidence intervals around μ and ν .

5.2. Elicitation mechanisms

The belief elicitation tasks that we use in Step 2 of Mechanism 1 provide incentives that are summarized, for each state θ_k , by a function

$$\pi_k : [0, 1] \rightarrow \mathcal{L}_X^A, \quad (8)$$

which takes as input the agent’s reported belief $r_k \in [0, 1]$, and returns as output the action-contingent lottery $\pi_k(r_k) \in \mathcal{L}_X^A$ that the agent will receive in return. Incentive-compatibility says that the agent’s actual report p_k will be the unique solution to the following maximization problem

$$\max \left\{ \underbrace{(1 - \beta_k)u_k(\pi_k(r_k)(a_\nu)) + \beta_k u_k(\pi_k(r_k)(a_\mu))}_{\text{SEU from reporting } r_k \text{ under utility } u_k \text{ and belief } \beta_k} \mid r_k \in [0, 1] \right\}.$$

Given that the utility u_k does not vary across A , there is a wide range of incentive-compatible elicitation mechanisms, e.g., proper binarized scoring rules (Hossain and Okui, 2013; Schlag and van der Weele, 2013), the Karni mechanisms (Karni, 2009; Tsakas, 2019), and mechanisms based on the method of matching probabilities (Ducharme and Donnell, 1973; Kadane and Winkler, 1988; Baillon et al., 2018) would all yield the true belief β_k as the unique optimal report irrespective of the shape of u_k . The common denominator among these mechanisms is that they are all binarized, i.e., they pay in lotteries over two fixed outcomes, a good and a bad one. Hence, theoretically, they would all guarantee truth-telling irrespective of the risk-attitudes that are induced by u_k . This is in contrast

¹⁵In practice, calibration of the updating parameter is typically done by comparing the Bayesian posterior with the one that the agent has reported, in settings where the prior and the likelihoods are exogenously set. The most common such task is the well-known urn experiment, which is widely used in the lab. For an overview of the extensive experimental literature —both in economics and psychology— that aims at calibrating the inference and base-rate parameters, we refer once again to the recent review of Benjamin (2019).

¹⁶In their paper, Offerman et al. (2009) recognize that the belief elicitation (under state-independent utilities) using the standard quadratic scoring rules can in principle be biased due to non-linear risk-preferences and probability weighting. Nevertheless, they would like to keep using this mechanism due to its simplicity. But at the same time, using out of sample data, they calibrate risk-preference and probability-weighting parameters, which are sufficient in order to debias the elicited beliefs.

with non-binarized scoring rules, like for instance the widely-used quadratic scoring rule (Brier, 1950; Good, 1952; Savage, 1971), which theoretically would also require us to assume risk-neutrality.

However, from an empirical point of view, the issue is far from settled. For instance, there is some experimental evidence against binarized scoring rules (Selten et al., 1999), but this view is not unanimous (Harrison et al., 2013, 2014, 2015). Typically, the argument against binarized scoring rules relies on the fact that the agent will have to reduce compound acts into simple ones, which is not a trivial exercise. One approach to deal with this problem is to use the relatively simpler quadratic scoring rule and then debias it (Offerman et al., 2009).¹⁷ Another more radical approach would be to omit the explicit description of the underlying incentives to the subject (Danz et al., 2020).

In any case, we find it outside the scope of this paper to participate in this debate, as the issue is orthogonal to our approach. Actually, the analyst can choose the method that in their view elicits most accurately the agent’s posterior beliefs and insert it in Step 2 of our Mechanism 1. Crucially, this choice of an elicitation task will not interact with the rest of the mechanism. That is, as long as we are confident that beliefs about A are elicited truthfully, the identification problem of beliefs about Θ will be resolved.

5.3. Suitable influential actions

As we have already discussed, the only assumption that we cannot dispense under any circumstances is the existence of an influential action that does not enter directly the agent’s utility function. But even when such an influential actions exists—which in most applications it does—how feasible is it to use it? In order to answer this question, we should first identify the potential obstacles for using an influential action in practice.

The main practical issue that we foresee is that in some applications, certain influential actions are too costly to implement. For instance, in Example 1 a donation to a political campaign would incur high financial cost for the analyst, and the same holds true in Example 2 where investing on a company would require significant funds. Still, there are two ways to mitigate this problem. First, the same influential action can be used for many agents simultaneously, meaning that the cost per subject becomes substantially smaller. For instance, once again in Example 1, suppose that we want to elicit the beliefs of a large sample of voters, potentially including both Democrats and Republicans. In this case, we can simply randomize once over $A = \{\text{donation, no donation}\}$, and use the same realization in the implementation of everyone’s mechanism. The second way to mitigate the problem is to reduce the cost of the influential action in expectation, by increasing the probability p of the default action. Of course, this is always feasible, as p is the analyst’s choice variable. Last but not least, one can always switch to an influential action of the second type, which is typically cheaper. Again in the context of our example with the Democrat, one could imagine presenting the agent with some poll results (like in the continuation of Example 1), which are true with some probability and fake otherwise. This would be certainly cheaper to implement.

The other potential problem is that in some occasions influential actions may be unethical. For instance in the context of Example 4, using data falsification in order to figure out your employee’s beliefs cannot be justified in any reasonable way as ethical conduct. Of course, such influential actions should never be used in practice. At the same time though, the range of influential actions is usually large, meaning that an alternative, less questionable influential action can often be found.

¹⁷In fact this method takes care both of distortions due to non-neutral risk preferences, as well as of distortions due to probability weighting.

6. Application: identifying motivated beliefs

It is widely accepted in the literature that people hold *motivated beliefs*, i.e., they systematically report biased beliefs in the direction of the state they prefer (Kunda, 1990).¹⁸ For instance, Democrats and Republicans systematically disagree on the probability they reportedly assign to some politically charged event, e.g., the winner of the next elections (Bullock et al., 2015).¹⁹ The literature on politically motivated beliefs would say that the two groups actually have different beliefs. However, this approach is in principle very simplistic in at least two ways. Let us explain where our concerns come from, and also argue that our mechanism can actually accommodate these concerns.

6.1. Misreporting

First of all, there is an alternative explanation: actual beliefs are not as divergent as the self-reported data suggest, and differences are amplified by misreporting (in the direction of their preferred parties respectively). Being able to test this alternative hypothesis is often crucial for policy purposes. We argue that our mechanism can do so.

For starters, note that misreporting of beliefs can be due to different psychological factor, e.g., 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). But it could also be potentially attributed to the subjects’ rational response to some (incentive-compatible) belief elicitation task due to the utility functions being state-dependent. The latter seems plausible, given that utilities will most likely be state-dependent in an environment with strong preferences over the state space. Actually, it is not difficult to see that our Mechanism 1 should definitely circumvent misreporting in the latter case (where misreporting is a rational response to the incentives). But in fact, we are also confident that it could also mitigate misreporting even in the former cases (where misreporting is attributed to psychological factors), as the subjects will not have direct preferences over A , which is the state space over which they eventually report probabilities. As a result, none of the aforementioned reporting biases should arise. This last conjecture should be seen as an interesting open question to be tested empirically, which is why we present it just as a plausible hypothesis.

6.2. Motivated reasoning

In most empirical studies, the conclusion that there are motivated beliefs is based on measurements of beliefs at a single instance, and subsequently on comparison of the empirical distributions of beliefs across different groups (Svenson, 1981). Although this approach is simple, and a fortiori experimentally convenient, it does not allow us to identify motivated beliefs at the individual level. As a result, it is nearly impossible to attribute motivated beliefs to the way different individuals internalize information. For instance, a common hypothesis is that Democrats will take into account a poll which is favorable to the Democratic candidate more than Republicans would. This phenomenon is called *motivated reasoning*.

To address this shortcoming of the early literature, recent research has focused on sophisticated experimental designs that elicit each subject’s individual beliefs at multiple instances, so that we can directly observe how different people update their beliefs in response to different information (Zimmermann, 2020; Thaler, 2020). However, these are involved experiments that aim at establishing and explaining motivated reasoning in specific settings.

¹⁸For a relatively recent review on the subject, see Benabou (2015).

¹⁹Based on data that we collected for a different project before the last American presidential elections, the average reported probability of Trump winning was 65% among Republicans and 40% among Democrats.

Our Mechanism 1 also addresses the same issue, but it does so with a simple task that can easily be used in a very broad range of settings. In particular, consider an influential action of the second type, e.g., agents are presented with a poll suggesting a landslide victory for the Democratic candidate, and they are told that the poll is true with some known probability and fabricated otherwise. Then, our Mechanism can identify both the benchmark belief μ and the updated belief ν given the poll for each individual agent. This means, that we can pin down the information that each individual actually internalizes upon seeing this poll, and therefore we can compare said internalized information for individual Democrats and individual Republicans. In other words, we can test motivated reasoning.

7. Final remarks

7.1. Null states

Throughout the paper we have assumed that both μ and ν are full-support beliefs, i.e., $\mu, \nu \in (0, 1)$, which in turn implies $p_0, p_1 \in (0, 1)$ for every $c > 0$ and $d > 0$. As it turns out, this assumption is not essential, and all our results still hold even if we relax it. Suppose that we only maintain the assumption that a_ν changes the agent's beliefs, i.e., $\mu \neq \nu$. This would imply $p_0 \neq p_1$. Then, it is not difficult to verify that the agent's beliefs will be identified by the following table:

	$p_1 = 0$	$0 < p_1 < 1$	$p_1 = 1$
$p_0 = 0$	$\mu = \nu$	$\mu = 1$ $\nu = \frac{1}{\alpha_1}$	$\mu = 1$ $\nu = 0$
$0 < p_0 < 1$	$\mu = 0$ $\nu = \frac{1}{1-\alpha_0}$	Theorem 2	$\mu = 1 - \alpha_0$ $\nu = 0$
$p_0 = 1$	$\mu = 0$ $\nu = 1$	$\mu = \alpha_1$ $\nu = 1$	$\mu = \nu$

Of course, if we want to assume Bayesian updating, all that is needed is to set $c = d = 1$ in the previous table, and replace Theorem 2 with Theorem 1 in the middle cell.

7.2. Identifying utility functions

Our identification approach focuses only on beliefs and does not tell us anything about the agent's utility function at each Θ -state. This is because we do not collect traditional choice data. However, as it is well-known in the literature, once we identify the belief μ , it is rather easy to also identify the exact marginal utilities at each state. To do so, it would suffice to observe traditional choices over acts. This is probably obvious to the trained decision-theorist, but let us anyway briefly outline one way to do this. First of all, notice that the set of all SEU representations can be obtained from preferences over acts, and it contains all pairs $(\tilde{u}, \tilde{\mu})$ such that

$$\tilde{u}_0 = a + b \left(k + \frac{1 - \mu}{1 - \tilde{\mu}} u_0 \right) \quad \text{and} \quad \tilde{u}_1 = a + b \left(\frac{\tilde{\mu} - 1}{\tilde{\mu}} k + \frac{\mu}{\tilde{\mu}} u_1 \right),$$

for any $a, k \in \mathbb{R}$ and $b > 0$. Observe that this is a generalization of the earlier rescaling in (2), which only gave us a subset of the SEU representations. In any case, the important point is that once μ is

known (via our Theorem 1), the set of utility function becomes

$$\tilde{u}_0 = \tilde{a}_0 + bu_0 \quad \text{and} \quad \tilde{u}_1 = \tilde{a}_1 + bu_1,$$

where $\tilde{a}_0 := a + bk$ and $\tilde{a}_1 = a - bk(1 - \mu)/\mu$. This means that we can still shift utilities up and down by constants (in opposite directions) or by the same constant in the same direction, but we cannot multiply the state utilities by different constant factors. This implies that the relative marginal utility between the two states is now uniquely identified. And this is exactly what we typically care about, i.e., pinning down the marginal effect from shifting payoffs from one state to the other.

7.3. Subjective expected utility

So far in our entire analysis we have taken for granted that the agent is a SEU maximizer. However, this assumption can be relaxed, as in Mechanism 1 we can elicit beliefs about A using binarized tasks in an environment with state-independent utilities (over A). In particular, we can replace SEU with probabilistic sophistication (Machina and Schmeidler, 1992). Note here that the agent is still required to update her beliefs in a Bayesian manner without being a SEU maximizer. However, this is not a big issue as the two problems can be disentangled once we depart from a SEU framework (Machina and Schmeidler, 1995). In this sense, our identification strategy becomes even more appealing for practical elicitation purposes, as it depends little on underlying assumptions regarding the agent's preferences over acts.

7.4. Finite state space

So far we have only dealt with binary state spaces. It is not difficult to verify that our method can be extended to any a finite state space. This can be done either by repeatedly applying Mechanism 1 for different binary partitions of Θ , or by extending the mechanism to accommodate more states. One way or another, the crucial part is that, in doing so we would need at least $|\Theta| - 1$ influential actions. And of course, different influential actions should yield different beliefs on Θ . To see why this is the case, suppose that there exists a third state θ_2 , but we still have only one influential action. Then, we will have four unknowns, viz., the probabilities attached to θ_1 and θ_2 by μ and ν respectively. At the same time, we will only have three equations, viz., one corresponding to the report p_0 , one to p_1 and another one to p_2 . So, we cannot solve the system to identify μ . But, by adding one more influential action, we increase the number of equations, which makes identification again feasible.

A. Proofs

Proof of Theorem 1. The result follows directly by setting $c = d = 1$ in Theorem 2, which is proven below. \square

Proof of Theorem 2. Using Equation (6), we obtain

$$\frac{\beta_0}{1 - \beta_0} = \frac{p^d(1 - \mu)^c}{(1 - p)^d(1 - \nu)^c} \quad \text{and} \quad \frac{\beta_1}{1 - \beta_1} = \frac{p^d\mu^c}{(1 - p)^d\nu^c}. \quad (\text{A.1})$$

given θ_0 and θ_1 respectively. Conditional on state θ_k , the agent's utility function over X is u_k , irrespective of the realized action in A . This means that an incentive-compatible elicitation mechanism makes it uniquely rational for the agent to report truthfully, i.e., we will have $p_k = \beta_k$ for each

$k \in \{0, 1\}$. As a result, we can replace the reports p_0 and p_1 into the respective equations in (A.1), to obtain

$$\frac{1 - \mu}{1 - \nu} = \underbrace{\left(\frac{p_0}{1 - p_0}\right)^{1/c} \left(\frac{p}{1 - p}\right)^{d/c}}_{\alpha_0} \quad \text{and} \quad \frac{\mu}{\nu} = \underbrace{\left(\frac{p_1}{1 - p_1}\right)^{1/c} \left(\frac{p}{1 - p}\right)^{d/c}}_{\alpha_1}. \quad (\text{A.2})$$

Note that both α_0 and α_1 are well-defined: by $\mu, \nu \in (0, 1)$, it follows that $p_0, p_1 \in (0, 1)$. Manipulating the two equations, we obtain the following system of linear equations:

$$\begin{bmatrix} 1 & -\alpha_0 \\ 1 & -\alpha_1 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ \nu \end{bmatrix} = \begin{bmatrix} 1 - \alpha_0 \\ 0 \end{bmatrix}. \quad (\text{A.3})$$

It is not difficult to verify that the system will have a unique solution if and only if $\alpha_0 \neq \alpha_1$, which is the case if and only if $\mu \neq \nu$. The latter follows directly from a_ν being an influential action. Hence, we solve the system, and we obtain

$$\begin{bmatrix} \mu \\ \nu \end{bmatrix} = \frac{1}{\alpha_0 - \alpha_1} \cdot \begin{bmatrix} -\alpha_1 & \alpha_0 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 - \alpha_0 \\ 0 \end{bmatrix} = \frac{1 - \alpha_0}{\alpha_1 - \alpha_0} \cdot \begin{bmatrix} \alpha_1 \\ 1 \end{bmatrix}, \quad (\text{A.4})$$

which concludes the proof. □

References

- ANDREONI, J. (1991). Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime? *RAND Journal of Economics* 22, 385–395.
- ANSCOMBE, F.J. & AUMANN, R.J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics* 34, 199–205.
- AUMANN, R.J. & SAVAGE, L. (1971). Letter to Leonard Savage, 8 January 1971. *Essays on Economic Decisions under Uncertainty* 76–78 (by J. Dreze, 1987).
- BACCELLI, J. (2017). Do bets reveal beliefs? A unified perspective on state-dependent utility issues. *Synthese* 194, 3393–3419.
- (2020). The problem of state-dependent utility: a reappraisal. *British Journal for the Philosophy of Science* 72, Nr. 2.
- (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.
- BENABOU, R. (2015). The economics of motivated beliefs. *Revue d’Economie Politique* 125, 665–685.
- BENJAMIN, D. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics*, Ch.2, 69–186.
- 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.

- DANZ, D., VESTERLUND, L. & WILSON, A. (2020). Belief elicitation: limiting truth telling with information on incentives. *Working Paper*.
- DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7, 1–68.
- (1974). *Theory of probability*. Vol. 1, Wiley.
- DRÈZE, J. (1961). Les fondements logiques de l'utilité cardinale et de la probabilité subjective. *La Decision, Colloques Internationaux du CNRS (Paris)* 73–87.
- (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 the European Economic Association* 13, 363–380.
- FEDDERSEN, T. & PESENDORFER, W. (1998). Convicting the innocent: the inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review* 92, 23–35.
- FISHBURN, P. (1973). A mixture-set axiomatization of conditional subjective expected utility. *Econometrica* 41, 1–25.
- GRETHER, D. (1980). Bayes rule as a descriptive model: the representativeness heuristic. *Quarterly Journal of Economics* 95, 537–557.
- 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., MARTÍNEZ-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., MARTÍNEZ-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.
- (2011a). A theory of Bayesian decision making with action-dependent subjective probabilities. *Economic Theory* 48, 125–146.
- (2011b). Subjective probabilities on a state space. *American Economic Journal: Microeconomics* 3, 172–185.
- (2016). Axiomatic foundations of expected utility and subjective probability. *Handbook of the Economics of Risk and Uncertainty*, Vol. 1, 1–39.
- KARNI, E. & SAFRA, Z. (1995). The impossibility of experimental elicitation of subjective probabilities. *Theory and Decision* 38, 313–320.
- KARNI, E. & SCHMEIDLER, D. (2016). 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.
- KUNDA, Z. (1990). The case of motivated reasoning. *Psychological Bulletin* 108, 480–498.
- LU, J. (2019). Bayesian identification: a theory for state-dependent utilities. *American Economic Review* 109, 3192–3228.
- MACHINA, M. & SCHMEIDLER, D. (1992). A more robust definition of subjective probability. *Econometrica* 60, 745–780.
- (1995). Bayes without Bernoulli: simple conditions for probabilistically sophisticated choice. *Journal of Economic Theory* 67, 106–128.
- 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.
- PAUKER, S.G. & KASSIRER, J.P. (1975). Therapeutic decision making: a cost-benefit analysis. *New England Journal of Medicine* 293, 229–234.
- (1980). The threshold approach to clinical decision making. *New England Journal of Medicine* 302, 1109–1117.
- RAMSEY, F. (1931). Truth and probability. *The Foundations of Mathematics and Other Logical Essays*, Wiley, NY.
- 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.
- SCHLAG, K. & VAN DER WEELE, J. (2013). Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality. *Theoretical Economics Letters* 3, 38–42.
- 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. (2020). The fake news effect: experimentally identifying motivated reasoning using trust in news. *Working Paper*.
- (2021). The supply of motivated beliefs. *Working Paper*.
- TSAKAS, E. (2017). Reasonable doubt revisited. *Working Paper*.
- (2019). Obvious belief elicitation. *Games and Economic Behavior* 118, 374–381.
- (2020). Identification of misreported beliefs. *Working Paper*.
- WAKKER, P. (1989). *Additive representations of preferences*. Kluwer Academic Publishers.
- (1993). Unbounded utility for Savage’s “foundations of statistics”, and other models. *Mathematics of Operations Research* 18, 446–485.
- ZIMMERMANN, F. (2020). The dynamics of motivated beliefs. *American Economic Review* 110, 337–363.