

# Robust scoring rules

Citation for published version (APA):

Tsakas, E. (2018). *Robust scoring rules*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 023 <https://doi.org/10.26481/umagsb.2018023>

## Document status and date:

Published: 08/10/2018

## DOI:

[10.26481/umagsb.2018023](https://doi.org/10.26481/umagsb.2018023)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Elias Tsakas

**Robust scoring rules**

RM/18/023

**GSBE**

Maastricht University School of Business and Economics  
Graduate School of Business and Economics

P.O. Box 616  
NL- 6200 MD Maastricht  
The Netherlands

# Robust scoring rules\*

ELIAS TSAKAS<sup>†</sup>

*Maastricht University*

First Draft: July 2017

This draft: October, 2018

## Abstract

We study elicitation of latent (prior) beliefs when the agent can acquire information via a costly attention strategy. We introduce a mechanism that simultaneously makes it strictly dominant to (a) not acquire any information, and (b) report truthfully. We call such a mechanism a *robust scoring rule*. Robust scoring rules are important for different reasons. Theoretically, they are crucial both for establishing that decision-theoretic models under uncertainty are testable. From an applied point of view, they are needed for eliciting unbiased estimates of population beliefs. We prove that a robust scoring rule exists under mild axioms on the attention costs. These axioms are shown to characterize the class of posterior-separable cost functions. Our existence proof is constructive, thus identifying an entire class of robust scoring rules. Subsequently, we show that we can arbitrarily approximate the agent’s prior beliefs with a quadratic scoring rule. The same holds true for a discrete scoring rule. Finally, we show that the prior beliefs can be approximated, even when we are uncertain about the exact specification of the agent’s attention costs.

KEYWORDS: Belief elicitation, prior beliefs, rational inattention, hidden information costs, posterior-separability, Shannon entropy, population beliefs, testing decision-theoretic models.

JEL CODES: C91, D81, D82, D83, D87.

## 1. Introduction

Subjective beliefs constitute one of the most common latent variables of interest in economics (e.g., [Manski, 2004](#)). Having recognized this, statisticians and economists have developed mechanisms, called (*proper*) *scoring rules*, that incentivize the economic agent to reveal his true latent belief, irrespective of which this belief is. Due to their solid theoretical foundations (i.e., the fact that they

---

\*This paper previously circulated under the title “Eliciting prior beliefs”. I am greatly indebted to Antonio Penta, John Rehbeck, Burkhard Schipper and Peter Wakker for their valuable comments at different stages of this project. I would also like to thank Chris Chambers, Madhav Chandrasekher, Paul Heidhues, Peter Katuščák, Dorothea Kübler, Fabio Maccheroni, Massimo Marinacci, Marcus Pivato, Marciano Siniscalchi, Mathias Staudigl, Stefan Terstiege, Nikolas Tsakas, Mark Voorneveld and the audiences in EEA-ESEM (Cologne), LOFT (Bocconi), RWTH Aachen University and University of Athens for helpful comments and fruitful discussions. I am also grateful to Lars Wittrock for his research assistance. Finally I would like to thank the Economics Department at UC Davis for its hospitality while working on this project.

<sup>†</sup>Department of Economics (AE1), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; Homepage: [www.elias-tsakas.com](http://www.elias-tsakas.com); E-mail: [e.tsakas@maastrichtuniversity.nl](mailto:e.tsakas@maastrichtuniversity.nl)

are incentive-compatible), proper scoring rules have been extensively used both in the lab and the field, as well as in various applications.

One of the main concerns with scoring rules is that (the incentives provided by) the mechanism itself may affect the very same beliefs it tries to elicit. As [Schotter and Trevino \(2014, p.109\)](#) eloquently put it,

*“the very act of belief elicitation may change the beliefs of subjects from their true latent beliefs or the beliefs they would hold (respond to) if those beliefs were not elicited (we might have a type of Heisenberg problem)”.*

This problem is particularly important both from a theoretical and an applied point of view.

For starters, we naturally ask whether it is even theoretically possible to construct non-invasive incentive-compatible mechanisms for belief elicitation. Specifically in the context of choice under uncertainty, a negative answer would immediately raise foundational questions about the testability of theories that use subjective beliefs as an explanatory variable of behavior. Consequently, we would not be able to identify the decision rule used by the individual who chooses under uncertainty ([Manski, 2004](#)). Take, for instance, subjective expected utility as the benchmark model and ask whether the observed choices can be rationalized by this model. Obviously, in order to answer this question, we need to be able to measure the individual’s subjective beliefs at the moment of his decision, and in order to do so we need a mechanism that does not affect these beliefs. We further elaborate in [Section 6.1](#).

Second, from an applied point of view, non-invasive belief elicitation mechanisms are particularly useful for surveys whose aim is to elicit the distribution of beliefs in a population. For instance, consider a marketing campaign interested in eliciting the average subjective belief in a population of consumers (e.g., about a new product being superior to the existing ones), or a political campaign interested in the median belief in a population of voters (e.g., about a proposed project being successful). The common denominator in the previous two examples is that we are not interested in learning the true state of nature per se, but rather in finding out what the population believes about the state of nature. Thus, we draw a representative sample from the respective population, we elicit individual beliefs from the sample, and we use the empirical frequency to estimate the distribution of the population beliefs. Crucially, we want the individuals in the sample to report the beliefs they would have held, had the survey not taken place. Otherwise, our estimate of the population beliefs will be biased. We further elaborate on this type of incentivized surveys in [Section 6.2](#).

Overall, the aim of this paper is to construct scoring rules that elicit the individual’s prior beliefs (viz., the ones he would have held in the absence of our elicitation procedure), rather than some updated beliefs that he may form after paying extra attention to the problem in an attempt to exploit the incentives provided by the scoring rule.

We consider scoring rules in a model with hidden information costs, which typically emerge as an expression of rationally inattentive preferences à la [Sims \(2003\)](#) (for an overview, see [Caplin, 2016](#)). In our formal model, there is a (male) agent – henceforth called the subject – who has a latent (prior) probabilistic belief for some fixed event.<sup>1</sup> A (female) agent – henceforth called the experimenter – wants to elicit this belief, and to this end she asks the agent to report it. In order to incentivize him to report truthfully, she designs a scoring rule that rewards the subject on the basis of his report and the realization of the event. Before stating his report, the agent can acquire information through a *costly attention strategy* and then reports his belief after having perhaps updated his prior (see [Figure 1](#) for the timeline).

---

<sup>1</sup>Throughout the paper – being aligned with the rational inattention literature – the term “prior belief” refers to the belief held by the subject in the absence of elicitation.

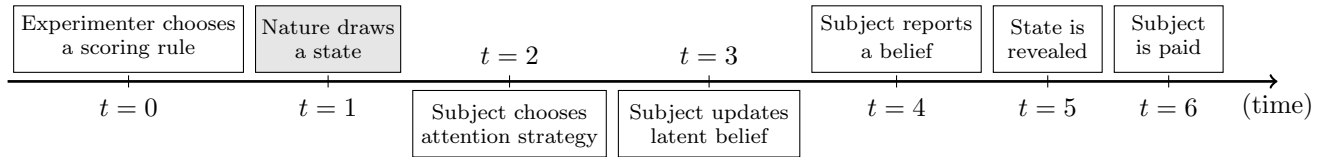


Figure 1: Boxes above the line are observed symmetrically by the subject and the experimenter. Boxes below the line are only observed by the subject. The shaded box is observed with a delay, i.e., it is realized at  $t = 1$  and observed at  $t = 5$ .

In order to elicit the subject’s prior belief, the scoring rule must make it simultaneously (a) strictly dominant not to acquire any information (i.e., to choose the degenerate zero-attention strategy), and (b) strictly dominant to report truthfully (i.e., the scoring rule is proper). Such a mechanism is called *robust scoring rule*. Two natural questions arise then. *Does a robust scoring rule exist?* And if yes, *how does it look like?* Note that, in expectation, every attention strategy yields a benefit (due to the fact that reporting is postponed till after the beliefs have been updated) and a cost (due to the fact that information acquisition is costly). Thus, the experimenter’s problem boils down to finding a scoring rule that provides sufficiently strong incentives for the agent to report truthfully, but not so strong to offset the costs of acquiring information. In this sense our work can be seen as part of a larger literature on the tradeoff between material payoffs and cognitive costs (Alaoui and Penta, 2018). The novelty of our work is that we try to exploit the presence of such costs.

Our main theorem shows that robust scoring rules exist under some mild axioms on the attention costs (Theorem 1). First, as always assumed in the literature, we require that the only costless attention strategy is the one that carries no information. Second, we impose a dynamic consistency axiom, which states roughly that it costs the same in expectation (to the subject) acquiring information at once or step-by-step. It turns out that our axioms characterize a well-known class of cost functions that has recently attracted attention in the literature, viz., the class of posterior-separable attention costs. Posterior-separability has solid theoretical foundations (Caplin et al., 2017; Zhong, 2017) and is supported by recent experimental evidence (Dean and Neligh, 2017). The proof of our theorem is constructive. Notably, not only do we show existence, but we also explicitly identify an entire class of robust scoring rules for each posterior-separable cost function. In this sense, our theory has strong empirical content.

We subsequently weaken our notion of robustness, by attempting to elicit a belief sufficiently close to the subject’s prior instead of the exact prior. In general, there are two forces that may bring the report away from the prior beliefs, viz., updating to some posterior belief (at  $t = 3$  in Figure 1) as a result of paying attention (at  $t = 2$ ), and reporting a belief different than the posterior (at  $t = 4$ ) as a result of the scoring rule not being proper. Approximate robustness requires that, even when combined, the two forces should not have a major effect. More precisely, a scoring rule is called  $\varepsilon$ -robust if, for every prior belief, the two forces will certainly lead to a report not further than  $\varepsilon$  away from the prior. Requiring merely approximate robustness allows us access to commonly-used scoring rules which we would not have been able to use, had we required exact robustness instead. In particular, we show that quite generally under our aforementioned axioms on the attention costs, for any margin of error  $\varepsilon > 0$ , there exists an  $\varepsilon$ -robust quadratic scoring rule (Proposition 2) and an  $\varepsilon$ -robust discrete scoring rule (Proposition 3). The previous results demonstrate that we do not need to use some exotic scoring rule in order for our theory to have a bite; instead, we can rely on usual scoring rules that are commonly-used in most applications and practical settings.

Resorting to approximate robustness has a second advantage, viz., it allows us to elicit beliefs sufficiently close to the prior, *even when we are uncertain of the subject’s attention costs*. In particular, suppose that the experimenter does not have enough data to calibrate the subject’s actual

cost structure, and instead she has formed a probabilistic estimate over the set of possible cost specifications. Let her accordingly assign probability  $p$  to the event that the subject’s attention costs satisfy our axioms. Then, we show that for every  $\varepsilon > 0$ , there exists a scoring rule that elicits a belief at most  $\varepsilon$  away from the prior with probability arbitrarily close to  $p$  (Theorem 2). This result is important for practical purposes, and in particular for surveys that aim at eliciting distributions of beliefs in a population. Indeed, typically in such settings, the experimenter has at best probabilistic estimates of the underlying costs (for further discussion, see Section 6.2).

On top of the foundational questions that we address (regarding the testability of decision-theoretic models under uncertainty), this paper contributes to two different streams of literature, viz., belief elicitation via scoring rules and rational inattention. A quick non-exhaustive overview of the main directions that have been explored so far in these two literatures is relegated to Section 7.

Of particular interest is the relationship between our paper and the one of Chambers and Lambert (2017) in that they are among the handful of papers that study dynamic belief elicitation. To the best of our knowledge, the only other paper is the one by Karni (2017).<sup>2</sup> In their paper, Chambers and Lambert (2017) consider an agent who has a latent prior belief and receives new information over time based on an exogenously given dynamic process. Then, they construct a mechanism which makes it incentive-compatible for the agent to simultaneously reveal his prior, his anticipated information flow and his realized posteriors. The conceptual difference to our paper is that the agent does not strategically choose the process of his information flow (viz., the attention strategy in our terminology). Moreover, the two papers differ in the formal approaches that they employ, viz., as opposed to our paper, their mechanism does not rely on the usual subgradient characterization, but rather on a randomization technique originally introduced by Allais (1953). On the other hand, a major similarity is that both our paper and the one of Chambers and Lambert (2017) truthfully elicit the agent’s prior beliefs.

The other paper that is closely related to our work is the one by Clemen (2002), who also studies the possible effect of scoring rules in information acquisition. However, unlike our paper, his aim is not to preclude information acquisition, as scoring rules are primarily used as incentive schemes for experts.

In Section 2 we introduce our basic framework. In Section 3 we present our axioms, we state our main result and we study the special case with entropic attention costs. In Section 4 we introduce approximate robustness, and we study approximately robust quadratic and discrete scoring rules. In Section 5 we study approximately robust scoring rules in the presence of uncertainty about attention cost. In Section 6 we further elaborate on our motivation. Section 7 contains a discussion. All proofs are relegated to the Appendix.

## 2. Preliminaries

**Proper scoring rules.** Consider a binary state space  $\Omega = \{\omega_0, \omega_1\}$ . A risk-neutral (male) experimental subject has a latent subjective belief  $\mu_0 \in [0, 1]$  of  $\omega_0$  occurring, which is not observed by the (female) experimenter. The subject is asked to state  $\mu_0$  and reports some  $r \in [0, 1]$ , which is not necessarily equal to  $\mu_0$ . A *scoring rule* is a function

$$S : [0, 1] \times \Omega \rightarrow \mathbb{R},$$

chosen by the experimenter, which takes the subject’s report ( $r$ ) and the realized state ( $\omega$ ) as an input and returns a monetary payoff ( $S_r(\omega)$ ) as an output. In economics we sometimes consider

---

<sup>2</sup>I am indebted to Chris Chambers for pointing out these connections.

stochastic scoring rules where the subject is paid in probabilities of winning a fixed prize. Stochastic scoring rules are used to elicit the subject's belief for arbitrary risk attitudes. Our entire analysis is directly extended to stochastic scoring rules (Section 7), implying that our assumption of the subject being risk-neutral is without loss of generality and can therefore be dispensed with. In statistics on the other hand the image of  $S$  is often allowed to take values in  $\overline{\mathbb{R}} = [-\infty, \infty]$ , in order to deal with a common subdifferentiability issue that often appears with ordinary scoring rules like the ones described above. For a discussion on such scoring rules, see Section 7.

The subject is assumed to maximize Subjective Expected Utility, i.e., given the scoring rule ( $S$ ) and a belief ( $\mu$ ), he chooses the report ( $r$ ) that maximizes

$$\mathbb{E}_\mu(S_r) := \mu S_r(\omega_0) + (1 - \mu) S_r(\omega_1).$$

A scoring rule is called proper whenever it is strictly dominant for the subject to report his true latent belief, irrespective of what this belief is. Formally,  $S$  is a *proper scoring rule*, whenever

$$\mathbb{E}_\mu(S_\mu) > \mathbb{E}_\mu(S_r) \tag{1}$$

for every  $r \neq \mu$  and every  $\mu \in [0, 1]$  (Brier, 1950; Good, 1952). The most commonly used proper scoring rule is the quadratic, which is defined by  $S_r(\omega_0) := \alpha - \beta(1 - r)^2$  and  $S_r(\omega_1) := \alpha - \beta r^2$ , where  $\alpha \in \mathbb{R}$  and  $\beta > 0$ . For a review of the standard proper scoring rules, see Schlag et al. (2015, Section 2).

It is well-known that a proper scoring rule is characterized by a class of strictly convex functions (McCarthy, 1956; Savage, 1971). In particular, the scoring rule  $S$  is proper if and only if there exists a strictly convex and subdifferentiable function  $\phi : [0, 1] \rightarrow \mathbb{R}$  such that  $S_r$  is a subtangent line at  $r$  evaluated at 1 and 0 respectively. Formally, let  $\phi$  be such that, for each  $r \in [0, 1]$  there are  $a_r, b_r \in \mathbb{R}$  such that  $\phi(s) \geq a_r + b_r s$  for all  $s \in [0, 1]$ , with equality holding if and only if  $s = r$ . In this case,  $S_r(\omega_0) := a_r + b_r$  and  $S_r(\omega_1) := a_r$  is a proper scoring rule. Strict convexity of  $\phi$  guarantees that (1) holds, while subdifferentiability at the boundary guarantees that the subtangent of  $\phi$  is not vertical and therefore  $S$  is well-defined. This last condition can be dispensed with, if we allow  $S$  to take values in  $\overline{\mathbb{R}}$  instead of  $\mathbb{R}$ , as often done in statistics. Whenever  $S$  is characterized by  $\phi$ , we obtain

$$\phi(\mu) = \mathbb{E}_\mu(S_\mu),$$

i.e.,  $\phi(\mu)$  is the subject's expected utility when he reports truthfully. As an example, the quadratic scoring rule is characterized by the function  $\phi_\beta(\mu) = \alpha - \beta\mu(1 - \mu)$ . For an overview of the subgradient characterization of proper scoring rules, see Gneiting and Raftery (2007).

**Costly attention.** We now enrich the agent's preferences to allow for information acquisition by means of costly attention (Sims, 2003). An *attention strategy* is a cognitive/thought experiment (viz., a signal), designed by the subject in an attempt to form updated "better" subjective beliefs. Given his prior  $\mu_0$ , each attention strategy is identified by a (Bayes-plausible) distribution over posteriors, chosen from the set  $\Pi(\mu_0) := \{\pi \in \Delta([0, 1]) : \int_0^1 \mu d\pi = \mu_0\}$ . We define the degenerate zero-attention strategy,  $\hat{\mu}_0 \in \Pi(\mu_0)$ , that puts probability 1 to the prior  $\mu_0$ . For notation simplicity, we henceforth denote by  $\hat{\Pi}(\mu_0) := \Pi(\mu_0) \setminus \{\hat{\mu}_0\}$  the set of non-degenerate attention strategies. If  $\mu \in \{0, 1\}$  then  $\hat{\Pi}(\mu) = \emptyset$ . Given a prior  $\mu_0$  and a scoring rule  $\phi$ , the (expected) benefit of an attention strategy  $\pi \in \Pi(\mu_0)$  is equal to

$$B_\phi(\pi) := \langle \phi, \pi \rangle - \phi(\mu_0),$$

where  $\langle \phi, \pi \rangle := \mathbb{E}_\pi(\phi)$  denotes the usual inner product duality. Since  $\phi$  is strictly convex, we obtain  $B_\phi(\pi) \geq 0$ , with equality holding if and only if  $\pi = \hat{\mu}_0$ . That is, attention always has strictly positive

benefits when the scoring rule is proper. However, attention is also costly. In particular, there is a cost function,

$$C : \Delta([0, 1]) \rightarrow \mathbb{R}_+$$

assigning a non-negative cost to every attention strategy  $\pi \in \Pi(\mu)$  for every prior  $\mu \in [0, 1]$ . Obviously the cost does not depend on the scoring rule, but only on the attention strategy. Attention costs can be identified from choice data (Caplin and Dean, 2015; Chambers, Liu and Rehbeck, 2017). They are part of the standard axiomatic characterizations of rationally inattentive preferences (De Oliveira et al., 2017; Ellis, 2018). The common entropic cost function is discussed in Section 3.3.

**Cost-benefit analysis.** Following the rational inattention literature, given a prior  $\mu_0$  and a proper scoring rule  $\phi$ , the subject will choose an attention strategy in  $\Pi(\mu_0)$  that maximizes the value

$$V_\phi(\pi) := B_\phi(\pi) - C(\pi).$$

After (optimally) choosing some  $\pi$ , the subject will first update his beliefs to some – also latent – posterior  $\mu \in \text{supp}(\pi)$ , and then – as  $\phi$  is proper – he will truthfully report his posterior belief  $\mu$ . Therefore, in order to guarantee that the agent will report his prior belief  $\mu_0$ , it must be the case that  $\hat{\mu}_0$  is a strictly dominant attention strategy. Whenever this is the case for every prior, we say that the scoring rule is robust.

**Definition 1.** A scoring rule  $\phi$  is *robust*, whenever it is proper and satisfies

$$V_\phi(\hat{\mu}) > V_\phi(\pi) \tag{2}$$

for every  $\pi \in \hat{\Pi}(\mu)$  and every  $\mu \in [0, 1]$ . ◁

Then, we naturally ask: *is there a robust scoring rule? And if yes, how does it look like?*

### 3. Main result

As turns out, a robust scoring rule exists under some mild regularity conditions, imposed on the cost function. In what follows in this section, we prove existence constructively, thus identifying an entire family of robust scoring rules.

#### 3.1. Axioms

We begin with two standard conditions, postulating that the zero-attention strategy is costless, whereas every other attention strategy is costly. Formally:

( $C_1$ ) NORMALIZATION:  $C(\hat{\mu}) = 0$  for all  $\mu \in [0, 1]$ .

( $C_2$ ) ATTENTION IS COSTLY:  $C(\pi) > 0$  for all  $\pi \in \hat{\Pi}(\mu)$  and all  $\mu \in [0, 1]$ .

The crucial restriction imposed by ( $C_1$ ) is that every zero-attention strategy induces the same cost irrespective of the prior  $\mu$ . The fact that this cost is set equal to 0 is merely a normalization. Note that under ( $C_1$ ), a scoring rule is robust if and only if  $V_\phi(\pi) < 0$  for every  $\pi \in \hat{\Pi}(\mu)$  and every  $\mu \in [0, 1]$ . Then, ( $C_2$ ) postulates that new information is always costly, in the sense that the cost of being attentive is higher than the normalized cost of the zero-attention strategy. This last condition is necessary for the existence of a robust scoring rule: indeed, if there is some  $\pi \in \hat{\Pi}(\mu)$



with  $C(\pi) \leq C(\hat{\mu})$ , then for every strictly convex  $\phi$  we obtain  $V_\phi(\hat{\mu}) < V_\phi(\pi)$ , implying that (2) is violated, and therefore the subject will always update his prior belief.

Our next axiom is relatively new to the literature, postulating that the cost of an attention strategy is only a function of the distribution of posteriors, and not of the underlying process that yields this distribution. Formally:

(C<sub>3</sub>) DYNAMIC CONSISTENCY: If  $\sigma : [0, 1] \rightarrow \Delta([0, 1])$  satisfies  $\sigma(\mu) \in \Pi(\mu)$  for all  $\mu \in [0, 1]$ , then

$$C(\mathbb{E}_\pi(\sigma)) = C(\pi) + \mathbb{E}_\pi(C \circ \sigma) \tag{3}$$

for all  $\pi \in \Delta([0, 1])$ .

Intuitively, if the subject chooses a sequential attention strategy, according to which he first picks  $\pi$  (first-period attention strategy) and then conditional on observing some posterior  $\mu \in \text{supp}(\pi)$  he picks a new attention strategy  $\sigma(\mu)$  (second-period attention strategy), the total cost that he incurs is equal to the cost of his first-period strategy ( $C(\pi)$ ) plus the expected cost of his second-period strategies ( $\mathbb{E}_\pi(C \circ \sigma)$ ). The distribution of posteriors at the end of the second period is then  $\mathbb{E}_\pi(\sigma)$ . Dynamic consistency postulates that the cost of an attention strategy that directly yields this distribution of posteriors (viz.,  $C(\mathbb{E}_\pi(\sigma))$ ) is equal to the total cost of the aforementioned sequential attention strategy (viz.,  $C(\pi) + \mathbb{E}_\pi(C \circ \sigma)$ ). Our notion of dynamic consistency is similar in spirit to the one in the standard characterization of dynamic variational preferences (Maccheroni et al., 2006).<sup>3</sup>

Our two axioms, (C<sub>1</sub>) and (C<sub>3</sub>), impose some basic coherency on the costs across different priors, similarly to recent work on dynamic information acquisition (Hébert and Woodford, 2016; Morris and Strack, 2017; Zhong, 2017), going beyond the decision-theoretic models of rational inattention, which specify a cost function  $C : \Pi(\mu) \rightarrow \mathbb{R}_+$  for each prior  $\mu \in [0, 1]$  but remain silent on the relationship of the costs across the different priors (De Oliveira et al., 2017; Ellis, 2018). Finally, as we show later in the paper, cost functions that satisfy (C<sub>1</sub>) – (C<sub>3</sub>) are canonical in De Oliveira et al.’s (2017) sense, i.e., they satisfy Blackwell monotonicity and convexity (see Section 7).

### 3.2. Existence

The following result answers our first question affirmatively, for the rather large class of cost functions that satisfy our axioms.

**Theorem 1** (Main result). *If the cost function satisfies (C<sub>1</sub>) – (C<sub>3</sub>), there is a robust scoring rule.*

The overall idea behind the previous result is to find a scoring rule that provides strong enough incentives to induce truth-telling (viz.,  $\phi$  must be strictly convex), but not so strong that lead the subject to update his beliefs (viz.,  $\phi$  should not be “too convex”). The proof is constructive, thus allowing us not only to prove that a robust scoring rule exists, but also to identify its functional form. Let us sketch the main steps here, while the full proof is relegated to Appendix A.

SKETCH OF THE PROOF AND INTUITION. We begin with the following intermediate result, which provides a characterization of the cost functions that satisfy our axioms by means of a property that has recently attracted interest in the rational inattention literature (Caplin et al., 2017). A similar result has been proven by Zhong (2017) in a somewhat different context, relying on standard properties of mutual information (e.g., Cover and Thomas, 2006).

---

<sup>3</sup>Of course in their paper the interpretation is different in that costs are incurred by nature, rather than by the decision maker. Nevertheless their condition – similarly to ours – guarantees that that costs are time-consistent. I am indebted to Fabio Maccheroni and Massimo Marinacci for pointing out this connection to me.

**Lemma 1.** *The cost function satisfies  $(C_1) - (C_3)$  if and only if it satisfies:*

POSTERIOR-SEPARABILITY: *There is a strictly concave function  $K : [0, 1] \rightarrow \mathbb{R}$  such that*

$$C(\pi) = K(\mu) - \langle K, \pi \rangle \quad (4)$$

*for every  $\pi \in \Pi(\mu)$  and every  $\mu \in [0, 1]$ .*

One can interpret  $K(\mu)$  as the cost of the most informative attention strategy when the prior belief is  $\mu$ , i.e., it is the cost that the subject must incur in order to learn the true state with certainty. The curvature of  $K$  puts a bound on the incentives that the scoring rule provides. Loosely speaking, “ $\phi$  must be less convex than  $-K$ ”, i.e., *a proper scoring rule  $\phi$  is robust if and only if the function*

$$\psi := \phi + K \quad (5)$$

*is strictly concave.*<sup>4</sup> Therefore, we must focus entirely on proper scoring rules that satisfy this last property. The most obvious such candidate is

$$f := a - bK \quad (6)$$

which is obviously strictly convex for every  $a \in \mathbb{R}$  and every  $b \in (0, 1)$ . In fact, if  $K$  is subdifferentiable at the boundary of  $[0, 1]$ , so is  $f$ , and therefore we can simply set  $\phi := f$ .

So let us focus on the case where  $K$  is not subdifferentiable, and a fortiori  $f$  is not either. One such example is the entropic cost function that we study in detail in the next section.

**Lemma 2.** *Consider a strictly convex function  $f : [0, 1] \rightarrow \mathbb{R}$ . Then, there exists a strictly convex and subdifferentiable function  $g : [0, 1] \rightarrow \mathbb{R}$  such that  $f - g$  is convex.*

The previous result, takes  $f$  as a benchmark and introduces the strictly convex function  $g$  which provides weaker incentives than  $f$ , i.e., formally,  $B_g(\pi) \leq B_f(\pi)$  for every  $\pi \in \Delta([0, 1])$ . Therefore, since  $B_f(\pi) < C(\pi)$ , it will also be the case that  $B_g(\pi) < C(\pi)$ , thus guaranteeing that (2) will be satisfied by  $g$ . Finally, since  $g$  is subdifferentiable, we can set  $\phi := g$ , thus completing the proof of our theorem. Note that our proof of Lemma 2 is constructive, implying that not only do we prove existence, but we also identify an entire family of robust scoring rules.  $\square$

### 3.3. Entropic attention costs

The most common functional form of attention costs within the rational inattention literature is the *entropic cost* specification (Sims, 2003, 2006; Caplin et al., 2017), which among other nice properties, allows us to provide microeconomic foundations to the multinomial logit model (Matejka and McKay, 2015). Accordingly, the cost of an arbitrary  $\pi \in \Pi(\mu_0)$  is equal to

$$C_\kappa(\pi) = \kappa(H(\mu_0) - \langle H, \pi \rangle), \quad (7)$$

where  $H(\mu) = -\mu \log \mu - (1 - \mu) \log(1 - \mu)$  is the Shannon entropy (Shannon, 1948), and  $\kappa > 0$  is a multiplier parameter. It is straightforward to verify that  $C_\kappa$  is posterior-separable with  $K = \kappa H$  being the corresponding function whose expected decrease gives the cost of attention. Therefore,  $C_\kappa$  satisfies  $(C_1) - (C_3)$ , and by Theorem 1, there exists a robust scoring rule. Since entropic attention costs are widely-used in applications and empirical studies, we go a step further asking whether there is a robust quadratic scoring rule  $\phi_\beta(\mu) = \alpha - \beta\mu(1 - \mu)$ . The following result answers the previous question affirmatively.

<sup>4</sup>Indeed, observe that  $\max_{\pi \in \Pi(\mu)} V(\pi)$  coincides with the concave closure of  $\psi$  evaluated at  $\mu$ . Hence, by Kamenica and Gentzkow’s (2011) concavification method we can characterize the optimal attention strategy.

**Proposition 1.** For an entropic cost function  $C_\kappa$ , the scoring rule  $\phi_\beta$  is robust if and only if  $\beta \leq 2\kappa$ .

Note that only the parameter  $\beta$  is relevant for robustness. This is not surprising, given that the incentives of a scoring rule are measured in terms of its convexity, and the constant  $\alpha$  does not affect the degree of convexity of  $\phi_\beta$ , but rather it simply rescales the payments by adding a constant.

The proof of the previous result exploits the fact that  $H$  is twice differentiable. Indeed, the condition  $\beta \leq 2\kappa$  is equivalent to  $\psi_\beta := \phi_\beta + K$  being strictly concave, which – as we have already discussed in the previous section – is equivalent to  $\phi_\beta$  being robust.

## 4. Approximate robustness

Often times in practical applications, the experimenter accepts a small error in the elicitation of the subject’s prior beliefs, either because she cannot detect small differences (say due to restrictions in the experimental technology) or because she does not care about minor mistakes in the elicited beliefs. Formally, for a given small  $\varepsilon \geq 0$ , the experimenter aims at picking a scoring rule that elicits a belief not further than  $\varepsilon$  away from the subject’s prior. We will call such a scoring rule  $\varepsilon$ -robust, or generally approximately robust.

Such willingness to accept minor elicitation errors allows the experimenter to choose from a wider set of scoring rules. In particular, she can now also use weakly proper scoring rules that make the subject indifferent between certain reports thus often leading to misreporting. Intuitively, a scoring rule is weakly proper if reporting truthfully is one of the optimal reports, but not necessarily the only one. Formally,  $S$  is *weakly proper* whenever, for all  $\mu \in [0, 1]$ ,

$$\mathbb{E}_\mu(S_\mu) \geq \mathbb{E}_\mu(S_r) \tag{8}$$

for every  $r \in [0, 1]$ . Geometrically, the function  $\phi$  that characterizes the scoring rule  $S$  is (only) weakly convex. Indeed,  $\phi$  may be linear in some subintervals  $I \subseteq [0, 1]$ , implying that every report (at least) in the interior of  $I$  yields the same utility vector, i.e.,  $S_r = S_{r'}$  for every  $r, r' \in \text{int}(I)$ . Thus, for a weakly proper scoring rule  $\phi$ , we define the set of optimal reports

$$I_\phi(\mu) := \arg \max_{r \in [0, 1]} \mathbb{E}_\mu(S_r) \tag{9}$$

when the posterior is  $\mu$ . Geometrically,  $I_\phi(\mu)$  is the largest interval of  $\mu$  where  $\phi$  is linear. Obviously, the scoring rule is proper if and only if  $I_\phi(\mu) = \{\mu\}$  for every  $\mu \in [0, 1]$ . Then, we are ready to define approximate robustness.

**Definition 2.** A scoring rule  $\phi$  is  $\varepsilon$ -robust, if for all  $\mu \in [0, 1]$  and for all  $\nu \in \text{supp}(\pi)$  such that  $\pi \in \arg \max_{\rho \in \Pi(\mu)} V_\phi(\rho)$ , it is the case that

$$I_\phi(\nu) \subseteq B_\varepsilon(\mu), \tag{10}$$

where  $B_\varepsilon(\mu) := \{\nu \in [0, 1] : |\mu - \nu| \leq \varepsilon\}$  is the closed  $\varepsilon$ -neighborhood of  $\mu$ . ◁

Intuitively, every optimal attention strategy  $\pi \in \Pi(\mu)$  puts positive probability to posteriors which will lead to reports not further away than  $\varepsilon$  from the prior. That is, the two forces that bring the optimal report away from the prior (viz., linearity of the scoring rule and incentives to acquire information), even when combined, will not lead further than  $\varepsilon$  away from the prior belief. Clearly, when a scoring rule is proper, it is  $\varepsilon$ -robust too if every optimal attention strategy yields posteriors at most  $\varepsilon$  away from the prior, i.e., in this case, only the second force (viz., the incentives to acquire information) may lead the subject to misreport.

Approximate robustness often proves to be useful by exploiting the tradeoff between simplicity of the scoring rule and the margin of error. Let us illustrate some cases where approximate robustness allows us to simplify the scoring rule, thus making it easier to implement.

## 4.1. Quadratic scoring rules

The scoring rule that is most commonly used in practice is the quadratic. As we have already mentioned above, a quadratic scoring rule is characterized by the function  $\phi_\beta(\mu) = \alpha - \beta\mu(1 - \mu)$ . While quadratic scoring rules are robust in some cases (e.g., see Section 3.3), it is not always so.

**Example 1.** Consider the posterior-separable cost function which is characterized by  $K(\mu) = -\mu^3$ . Note that the second derivative of  $K$  is not bounded away from 0, viz., the cost function becomes arbitrarily flat (close to 0). Now consider a quadratic scoring rule  $\phi_\beta$ , and observe that its second derivative is bounded away from 0. Then, we note that  $\psi''_\beta(\mu) = 2\beta - 6\mu$ , implying that  $\psi_\beta = \phi_\beta + K$  is convex in  $[0, \beta/3]$ , as illustrated in Figure 2 below. Hence,  $\phi_\beta$  is not robust for any  $\beta > 0$ .  $\triangleleft$

Of course, in the previous example, a robust scoring rule exists (by Theorem 1). However, it is not quadratic. Now suppose that the experimenter is willing to accept a margin of error, as long as she can use a quadratic scoring rule, which is in general deemed simpler and easier to understand. Thus, we ask: *is there an approximately robust quadratic scoring rule instead?*

**Proposition 2.** *If the cost function satisfies  $(C_1) - (C_3)$  with the corresponding  $K$  being polynomial, then for every  $\varepsilon > 0$  there exists an  $\varepsilon$ -robust quadratic scoring rule.*

The previous result can be directly extended to any  $K$  with finitely many roots of the second derivative. Given that most cost functions that are typically considered in the literature satisfy this last condition, our result is quite generic. Therefore, for practical purposes, *quadratic scoring rules can effectively elicit the subject's prior with arbitrary precision*. Let us illustrate our result in the context of our previous example.

**Example 1 (continued).** For starters, notice that for  $\beta \geq 3$ , the function  $\psi_\beta$  is strictly convex in  $[0, 1]$ , implying that for every prior the subject will choose the most informative attention strategy, and therefore  $\phi_\beta$  is not  $\varepsilon$ -robust for any  $\varepsilon < 1$ . Thus, we focus on quadratic scoring rules with  $\beta < 3$ . In each such case, as we have already shown above,  $\psi_\beta$  is strictly convex in  $[0, \beta/3]$  and strictly concave in  $[\beta/3, 1]$ , as illustrated in the following figure. For starters, verify that  $\psi_\beta$  is

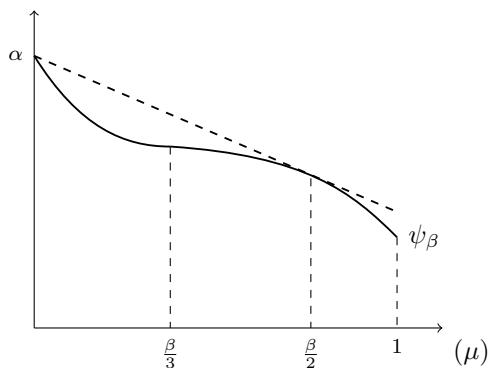


Figure 2: Approximately robust quadratic scoring rule.

strictly decreasing in  $[0, 1]$  for every  $\beta < 3$ . Then, note that there exists some  $\mu_\beta > \beta/3$  such that  $\psi'_\beta(\mu_\beta) = (\psi_\beta(\mu_\beta) - \psi_\beta(0))/\mu_\beta$ . Indeed, it turns out that  $\mu_\beta = \beta/2$  for every  $\beta < 3$ . That is, the concave closure of  $\psi_\beta$  is linear in  $[0, \beta/2]$  and strictly concave in  $[\beta/2, 1]$ . Hence, by [Kamenica and Gentzkow \(2011\)](#), if  $\mu \in (0, \beta/2)$  then the subject's optimal attention strategy is  $\pi \in \Pi(\mu)$  such that  $\text{supp}(\pi) = \{0, \beta/2\}$ , i.e., the subject will update his beliefs to either 0 or to  $\beta/2$ . On the other hand, if  $\mu \in [\beta/2, 1]$  then the subject's optimal attention strategy is  $\hat{\mu}$ , i.e., the subject will not update his prior. In both cases, the subject will report his posterior truthfully, as  $\phi_\beta$  is proper. Finally, for every  $\varepsilon > 0$ , if we set  $\beta := 2\varepsilon$  then  $\phi_\beta$  is  $\varepsilon$ -robust.  $\triangleleft$

The proof of Proposition 2 has a similar logic as the one of the previous example. In particular, for every  $\varepsilon > 0$ , we choose some  $\beta$  small enough so that the (finitely many) intervals where  $\psi_\beta$  is convex become small enough, in which case the subject does not use an attention strategy that brings him further than  $\varepsilon$  away from the prior.

## 4.2. Discrete scoring rules

Suppose that the subject can only report beliefs from the finite set

$$R := \left\{ \frac{0}{n}, \frac{1}{n}, \dots, \frac{n}{n} \right\}$$

for an arbitrary  $n \in \mathbb{N}$ .<sup>5</sup> We henceforth denote the  $k$ -th available report by  $r_k := k/n$  for each  $k \in \{0, 1, \dots, n\}$ . Such scoring rules are called discrete and are quite common in experiments where the subjects are asked to choose a report from some predetermined grid of probabilities. The practical advantage is that discrete scoring rules are easier to implement, as they can be presented to the subject in the form of a list. On the negative side, the experimenter can only hope for a  $(1/n)$ -approximation of the subject's beliefs, i.e., for each prior  $\mu \in [r_k, r_{k+1}]$  the experimenter will at best elicit  $r_k$  or  $r_{k+1}$ . Thus, we naturally ask: *Is it indeed always possible to construct a discrete scoring rule that elicits a belief at most  $1/n$  away from the subject's prior?* If yes, we then call it  *$(1/n)$ -robust discrete scoring rule*.

**Proposition 3.** *If the cost function satisfies  $(C_1) - (C_3)$ , there is an  $\frac{1}{n}$ -robust discrete scoring rule.*

The technical difficulty with discrete scoring rules is that they induce piecewise linear expected benefits, thus often providing stronger incentives for attention than the usual scoring rules which are strictly convex. This is particularly the case when the subject's latent belief lies nearby the kinks of the discrete scoring rule.<sup>6</sup> Hence, it becomes easier for a non-degenerate attention strategy to be profitable. Thus, we need to counteract these increased incentives, by making the scoring rule flatter, until we offset them. Let us provide some intuition on how we construct such a scoring rule for  $n = 2$ . The full proof (for every  $n \in \mathbb{N}$ ) is relegated to Appendix B.

**SKETCH OF THE PROOF (FOR  $n = 2$ ).** Fix a strictly concave function  $K$ . The set of available reports is  $R = \{0, 1/2, 1\}$ . For every pair of consecutive available reports, take the middle point, viz.,  $1/4$  and  $3/4$  respectively. Now, we consider a convex and piecewise linear function  $\phi$ , with kinks at the middle points,  $1/4$  and  $3/4$ . The extension of  $\phi$  from the interval  $[1/4, 3/4]$  to the entire unit interval, when evaluated at 0 and 1, yields the two possible payments when  $1/2$  is reported, and likewise for the other two intervals (viz.,  $[0, 1/4]$  and  $[3/4, 1]$ ) and the corresponding reports (viz., 0 and 1 respectively). Note that this discrete scoring rule elicits the closest report to the subject's (posterior) belief, as illustrated in Figure 3. For instance, for a posterior  $\mu \in [0, 1/2]$ , the subject reports 0 if  $\mu \in [0, 1/4]$  and reports  $1/2$  if  $\mu \in [1/4, 1/2]$ . Subsequently, observe that  $\phi$  is constructed in a way such that the concave closure of  $\psi = \phi + K$  is linear in two intervals  $[\mu_1^1, \mu_1^2] \subseteq [0, 1/2]$  and  $[\mu_2^1, \mu_2^2] \subseteq [1/2, 1]$ , and strictly concave in every other convex subset of  $[0, 1] \setminus ([\mu_1^1, \mu_1^2] \cup [\mu_2^1, \mu_2^2])$ . Therefore, by the usual concavification argument (Kamenica and Gentzkow, 2011), the subject will update his beliefs if and only if his prior belongs to  $(\mu_1^1, \mu_1^2) \cup (\mu_2^1, \mu_2^2)$ . Indeed, take for instance some  $\mu \in [0, 1/2]$ . Then consider the following cases: (i) if  $\mu \in (\mu_1^1, \mu_1^2)$  then the subject's optimal attention

<sup>5</sup>Our entire analysis can be directly extended to any finite set  $R \subset [0, 1]$ , e.g., in cases where the subject can only report a fraction of 2 or a fraction of 3.

<sup>6</sup>The kinks of the scoring rule do not necessarily coincide with the available reports. For instance in the upcoming example (Figure 3), the available reports are  $R = \{0, 1/2, 1\}$  whereas the kinks appear at  $\{1/4, 3/4\}$ .

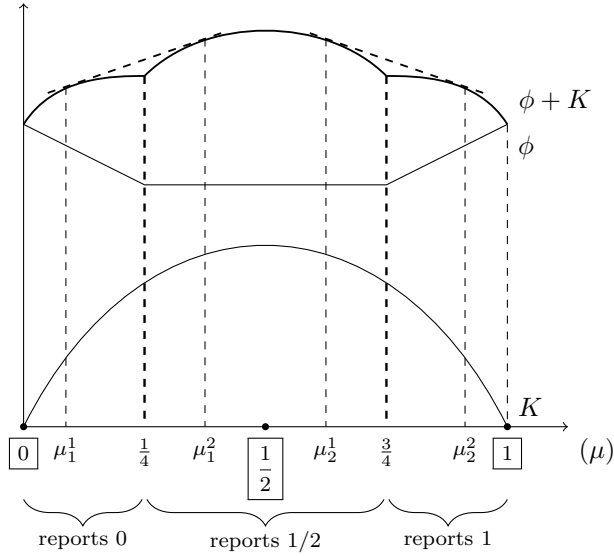


Figure 3: Approximately robust discrete scoring rules.

strategy is  $\pi \in \Pi(\mu)$  such that  $\text{supp}(\pi) = \{\mu_1^1, \mu_1^2\}$ , implying that she updates to either  $\mu_1^1$  (in which case he reports 0) or  $\mu_1^2$  (in which case he reports 1/2), (ii) if  $\mu \in [0, \mu_1^1]$  then his optimal attention strategy is  $\hat{\mu}$  and he reports 0, and (iii) if  $\mu \in [\mu_1^2, 1/2]$  then again his optimal attention strategy is  $\hat{\mu}$  and he reports 1/2. Hence, he will report either 0 or 1/2, implying that  $\phi$  is (1/2)-robust, as desired.  $\square$

## 5. Cost uncertainty

Sometimes, the experimenter cannot pin down the subject's cost function with certainty. This is typically due to the experimenter not having enough data to calibrate the subject's (actual) cost function. In such cases, she instead resorts to an estimated probability distribution over cost functions. *Can the experimenter then be sufficiently certain that she will approximate the subject's prior beliefs, although she is not certain of his cost specification?*<sup>7</sup>

Let  $\mathcal{C}$  denote the space of continuous functions  $K : [0, 1] \rightarrow \mathbb{R}$ , together with topology induced by the sup norm  $\|\cdot\|_\infty$ . Moreover, let  $\mathcal{K} \subset \mathcal{C}$  be the (convex) space of strictly concave functions such that  $K(0) = K(1) = 0$ . Each cost function satisfying  $(C_1) - (C_3)$  is identified by a unique  $K \in \mathcal{K}$ .<sup>8</sup> Outside  $\mathcal{K}$  we find costs  $K \in \mathcal{C}$  that violate one or more of our axioms  $(C_1) - (C_3)$ . For instance, a constant function  $K$  violates  $(C_2)$  by making every attention strategy costless, and as a result no scoring rule approximates the prior beliefs with precision better than 1, i.e., we get pure noise irrespective of the scoring rule that we use. Uncertainty about the subject's costs is described by a distribution  $P \in \Delta(\mathcal{C})$ . Whenever  $P(\mathcal{K}) = 1$ , the experimenter is certain that the subject's cost function satisfies  $(C_1) - (C_3)$ .

For each scoring rule  $\phi$ , define the function  $\varepsilon_\phi : \mathcal{C} \rightarrow \mathbb{R}_+$  by

$$\varepsilon_\phi^K := \inf\{\varepsilon \geq 0 : \phi \text{ is } \varepsilon\text{-robust given } K\}, \quad (11)$$

<sup>7</sup>I am indebted to Burkhard Schipper for suggesting this approach.

<sup>8</sup>In principle, each cost function is associated with an entire class of strictly concave functions. Our normalization identifies a single element of this class, and in this sense  $\mathcal{K}$  represents the set of costs that satisfy  $(C_1) - (C_3)$ . Note that focusing on continuous functions is without loss of generality: by concavity,  $K$  is continuous in  $(0, 1)$ , and possible discontinuities at the boundary make it easier for a scoring rule to be robust. We further elaborate in Footnote 10 in Appendix C.

which provides a bound on the approximation of the prior that the experimenter can achieve with  $\phi$  for each  $K \in \mathcal{C}$ . Note that  $\varepsilon_\phi$  is upper semi-continuous (Lemma C1). Hence,  $\{K \in \mathcal{C} : \varepsilon_\phi^K \leq \varepsilon\}$  is Borel measurable. That is, the event that “the scoring rule  $\phi$  is  $\varepsilon$ -robust” is expressible in the experimenter’s language, and therefore the experimenter assigns some probability to it.

**Definition 3.** A scoring rule  $\phi$  is  $(\varepsilon, \delta)$ -robust (or in general, approximately robust) if

$$P(\{K \in \mathcal{C} : \varepsilon_\phi^K \leq \varepsilon\}) \geq 1 - \delta, \quad (12)$$

given some fixed  $\varepsilon \geq 0$  and  $\delta \geq 0$ . ◁

That is, the probability of eliciting a belief further than  $\varepsilon$  away from the prior is smaller than  $\delta$ . *Can we always find an approximately robust scoring rule under uncertainty about the cost function?*

**Theorem 2.** *If  $P(\mathcal{K}) = p$ , then for all  $\varepsilon > 0$  and  $\delta > 1 - p$  there is an  $(\varepsilon, \delta)$ -robust scoring rule.*

The intuition of the previous result is that we can construct a scoring rule that  $\varepsilon$ -approximates the prior with probability arbitrarily close to the probability that the experimenter attaches to the cost function satisfying  $(C_1) - (C_3)$ . For instance, if the experimenter is certain that the subject’s cost satisfies  $(C_1) - (C_3)$ , then she can approximate the prior with probability arbitrarily close to 1.

SKETCH OF THE PROOF. The proof is somewhat similar in logic to the one of Proposition 3. In particular, for an arbitrary  $\varepsilon > 0$  we take some  $n \in \mathbb{N}$  such that  $2/n \leq \varepsilon$ . Then, we construct a weakly convex and piecewise linear function  $\phi^\gamma$  with kinks at every  $k/n$ , such that the slope between two consecutive linear segments increases by a constant term of  $\gamma > 0$ . Then, we show that for every  $K \in \mathcal{K}$  there exists some  $\gamma$  which is sufficiently small to guarantee that the concave closure of  $\psi^\gamma := \phi^\gamma + K$  will only be linear on intervals of length at most  $1/n$ . Hence, for an arbitrary prior  $\mu \in [0, 1]$ , the optimal attention strategy will yield a posterior at most  $1/n$  away from  $\mu$ , and the subject will subsequently report a belief at most  $1/n$  away from the posterior. Hence, overall the subject with a cost function  $K$  will report a belief not further than  $\varepsilon$  away from the prior, when the scoring rule is  $\phi^\gamma$ . Finally, we select  $\gamma$  to be sufficiently small so that sufficiently many cost functions satisfy the aforementioned property. More specifically, for any given  $\delta > 1 - P(\mathcal{K})$ , we show that there is a small enough  $\gamma$ , such that the set  $\{K \in \mathcal{K} : \phi^\gamma \text{ is } \varepsilon\text{-robust given } K\}$  receives probability at least equal to  $1 - \delta$ . ◻

Of course, the scoring rule proposed in our (constructive) existence proof is not the only one. Indeed, we can often find other  $(\varepsilon, \delta)$ -robust scoring rules. For starters, it is obvious that our result can be directly adjusted to discrete scoring rules. Furthermore, we can often obtain  $(\varepsilon, \delta)$ -robust quadratic scoring rules, as illustrated in the example below.

**Example 2.** Let  $P$  be a uniformly distributed over the set of entropic cost functions in  $\mathcal{E} := \{C_\kappa | \kappa \in [0, 1]\}$  with multiplier parameter in the unit interval (see Section 3.3). Obviously, the closer  $\kappa$  gets to 0, the more difficult it becomes for a scoring rule to be robust. Indeed, there is no quadratic scoring rule that elicits the subject’s prior with probability 1. Nevertheless, for any  $\delta > 0$  (arbitrarily close to 1), a quadratic scoring rule  $\phi_\beta$  with  $\beta \leq 2(1 - \delta)$  will elicit the subject’s prior with probability larger or equal than  $1 - \delta$  (by Proposition 1). ◁

## 6. Revisiting our motivation

### 6.1. Testing decision-theoretic models under uncertainty

Suppose that we want to test a decision-theoretic model under uncertainty. For the sake of our argument, consider Subjective Expected Utility (SEU) as our benchmark model, noting that we need to have data about both the individual’s choices and his (prior) beliefs. As we have already discussed throughout the paper, the term “prior beliefs” refers to his beliefs at the moment his choice is made. In fact, it is plausible that before encountering the decision problem, he may not even have well-formed beliefs. The latter is not deemed to be an issue for us, as we are not interested in how beliefs are formed in the period leading to the individual’s decision, but rather how he responds to these beliefs. In particular, *we reject SEU if and only if the observed choice does not maximize the individual’s SEU given his prior beliefs*. Hence, all our efforts are going to focus on the elicitation of the individual’s prior beliefs.

For starters, notice that elicitation must take place after the choice has been made, as we cannot control how the individual responds to the incentives provided by the decision problem itself. Note that, unlike the experimental literature, the problem here is not whether belief elicitation will affect the behavior,<sup>9</sup> but rather that the individual may use a (non-degenerate) attention strategy between the elicitation stage and the decision stage (as a response to the material payoffs provided by the decision problem). Hence, elicitation must necessarily follow the decision. In this case, the only incentives that the individual has are the ones that stem from the scoring rule. Therefore, as long as the scoring rule is robust, a rational individual will not further update his beliefs.

Finally, let us stress that our work is not concerned with the experimental mechanics of belief elicitation. In particular, we have nothing substantial to contribute to the debate on the effect of belief elicitation on behavior, or to the question of whether incentives matter empirically when eliciting beliefs, or to the question of whether experimental subjects best respond to their stated beliefs. What we are actually establishing is that it is theoretically feasible to test decision-theoretic models under uncertainty, something far from being *ex ante* obvious.

### 6.2. Eliciting population beliefs via surveys

Suppose that we want to elicit the distribution of beliefs about some event in a large population. This would be of particular relevance, for instance, to marketing practitioners who are interested in the public perception for some product or political campaigns who would like to know the public opinion on some proposed reform. In these cases, *we are not interested in obtaining an estimate of the true state of nature, but rather of what the population believes about the state of nature*.

To elicit such population beliefs we need to first truthfully elicit individual prior beliefs in a (large) representative sample, and then use the empirical frequency as an estimate of the population’s distribution. Note the importance of eliciting each individual’s *true prior beliefs*. Suppose, for instance, that the true distribution of beliefs is bell-shaped, and let the elicitation mechanism provide strong incentives so that a significant portion of the sample chooses informative attention strategies, which push their posterior beliefs towards the two extremes, thus leading to a bimodal empirical frequency with the corresponding biases. Furthermore, note that in such a setting individual cost functions are typically not known. Thus, our theory of approximate robustness under cost uncertainty (Section 5) becomes very relevant.

---

<sup>9</sup>Even experimentally there is no consensus as to whether belief elicitation affects followup decisions (Schotter and Trevino, 2014, Sect. 3.3).



As we have already mentioned, we are not interested in learning the true state of nature, but rather what the population believes about it, which is why standard elicitation mechanisms, such as prediction markets (Hanson, 2003), would most likely yield biased estimates, primarily because they typically attract participants who put significant effort into learning the true state. Finally, traditional surveys are typically not incentivized (for an exception to this rule, see Grisley and Kellogg, 1983). This is mainly due to practical reasons. For instance, providing monetary incentives is often very costly. Additionally, beliefs are mainly elicited as part of large household surveys and are typically about long-term events, which therefore cannot be incentivized based on their realization. Our proposed theory on the other hand primarily applies to surveys run by large firms or organizations or campaigns with stakes in the population beliefs about short-term events.

## 7. Discussion

**Other related literature.** Scoring rules were originally introduced by meteorologists (Brier, 1950), before being further developed by statisticians (Good, 1952; McCarthy, 1956; Savage, 1971), and eventually being adopted by several disciplines, such as economics, accounting, business, management, psychology, political science and computer science (see Offerman et al., 2009, p. 1462). Within economics, the theory of scoring rules has mostly focused on introducing new mechanisms (Hossain and Okui, 2013; Karni, 2009), on relaxing the underlying assumptions of the standard mechanisms, such as for instance risk-neutrality (Savage, 1971; Offerman et al., 2009; Schlag and van der Weele, 2013) or the expected utility hypothesis (Karni, 1999; Chambers, 2008; Offerman et al., 2009), and on understanding the technical relationships to other economics models (Chambers, Healy and Lambert, 2017). Scoring rules are also used in various economic applications, focusing for instance on incentive schemes in organizations (Thomson, 1979), information markets (Hanson, 2003; Ostrovsky, 2012) and strategic indistinguishability (Bergemann et al., 2017). Finally, there is large experimental literature, focusing primarily on the role of risk-aversion (Offerman et al., 2009; Armantier and Treich, 2013), the comparison of deterministic and stochastic scoring rules (Selten et al., 1999; Harrison et al., 2013, 2014) and the elicitation of beliefs in games (Nyarko and Schotter, 2002; Costa-Gomes and Weizsäcker, 2008; Palfrey and Wang, 2009; Rutström and Wilcox, 2009). For two recent literature reviews, we refer to Schotter and Trevino (2014) and Schlag et al. (2015).

Rational inattention models first appeared in macroeconomics (Sims, 2003, 2006; Maćkowiak and Wiederholt, 2009), before attracting interest of microtheorists. The latter have mostly focused on providing axiomatic foundations (De Oliveira et al., 2017; Ellis, 2018) and on designing revealed-preference tests for identifying the attention costs from choice data (Caplin and Dean, 2015; Chambers, Liu and Rehbeck, 2017; Caplin et al., 2017). Recently, there is interest in dynamic models of rational inattention (Hébert and Woodford, 2016; Morris and Strack, 2017; Zhong, 2017). There are also various economic applications of rational inattention – usually with entropic costs – on topics like discrimination (Bartoš et al., 2016), pricing (Matejka, 2016) and electoral competition (Matejka and Tabellini, 2016). Finally, there is recent work on experimentally testing models of rational inattention (Dean and Neligh, 2017). For an overview of this literature, see Caplin (2016).

**Other experimental currencies.** Throughout the paper we have focused exclusively on scoring rules that pay in monetary payoffs, i.e., formally,  $S$  takes values in  $\mathbb{R}$ . However, as we have already mentioned, there are large literatures dealing with scoring rules that pay either in probabilities over a fixed prize or allowing for infinite rewards/losses, i.e., formally,  $S$  takes values in  $[0, 1]$  in the former and in  $\mathbb{R}$  in the latter case, respectively. Let us first briefly present the motivation and then discuss our results for each of these alternative mechanisms.

Scoring rules that pay in probability currencies are called stochastic and have been introduced in

economics in order to deal with subjects who are not risk-neutral (Savage, 1971; Schlag and van der Weele, 2013). Formally,  $S_r(\omega) \in [0, 1]$  is the objective probability of the subject winning the prize, when he reports  $r$  and the realized state is  $\omega$ . In this case the subject's expected utility is linear in the probability of winning the prize irrespective of his risk preferences, and our analysis follows verbatim except for one small detail, viz., in order for a function  $\phi$  to characterize a stochastic scoring rule, not only should it be subdifferentiable, but it should also have at every point a subtangent that takes values in  $[0, 1]$  both when evaluated at 0 and at 1. The latter holds whenever  $\phi$  satisfies the following two conditions:  $0 \leq \phi(0) + \phi'(0) \leq 1$  and  $0 \leq \phi(1) - \phi'(1) \leq 1$ , for some  $\phi'(\mu) \in \partial\phi(\mu)$ , for each  $\mu \in \{0, 1\}$ . Hence, for a robust  $\phi$  there exists some  $\beta \in \mathbb{R}$  and a sufficiently small  $\gamma \in (0, 1)$  such that  $\psi(\mu) := \gamma(\phi(\mu) + \beta\mu)$  satisfies the previous two inequalities, thus implying that  $\psi$  is a robust stochastic scoring rule. It is important to mention that stochastic scoring rules have been criticized based on experimental evidence (Selten et al., 1999), although such criticism is not unanimous (Harrison et al., 2013, 2014). For an in-depth discussion on the role of risk preferences, we refer to Offerman et al. (2009).

Scoring rules that allow for infinite rewards/losses are common in statistics (Gneiting and Raftery, 2007). The main reason behind such generalization is in order to be able to dispense with the subdifferentiability of  $\phi$ . More specifically, when  $S$  is allowed to take infinite values, every strictly convex function  $\phi$  characterizes a proper scoring rule, even if it is not subdifferentiable at the boundary. In this last case, the respective subtangents are infinitely sloped, which is why  $S$  needs to be unbounded. In fact, under such generalized scoring rules, the proof of our main result becomes straightforward, viz., the function  $f = \gamma - \lambda K$  (see (6)) is always robust. However, it would still be very difficult to practically implement such a scoring rule.

**Canonical attention costs.** In the rational inattention literature, there are two natural regularity properties that we typically require the cost function to satisfy, viz., Blackwell monotonicity and convexity (De Oliveira et al., 2017). First, let us define the (partial) Blackwell order in  $\Pi(\mu)$ . For two attention strategies,  $\pi, \rho \in \Pi(\mu)$ , we say that  $\pi$  is Blackwell more informative than  $\rho$ , and we write  $\pi \succeq \rho$ , whenever  $\langle f, \pi \rangle \geq \langle f, \rho \rangle$  for every convex function  $f : [0, 1] \rightarrow \mathbb{R}$  (Blackwell, 1953). The two axioms postulate that, for every  $\mu \in [0, 1]$ , the following hold respectively:

(C<sub>4</sub>) BLACKWELL MONOTONICITY:  $C(\pi) \geq C(\rho)$  for all  $\pi, \rho \in \Pi(\mu)$  with  $\pi \succeq \rho$ .

(C<sub>5</sub>) CONVEXITY:  $C(\lambda\pi + (1 - \lambda)\rho) \leq \lambda C(\pi) + (1 - \lambda)C(\rho)$  for all  $\pi, \rho \in \Pi(\mu)$  and all  $\lambda \in (0, 1)$ .

The obvious question is whether our axioms (C<sub>1</sub>) – (C<sub>3</sub>) imply (C<sub>4</sub>) – (C<sub>5</sub>), i.e., are the cost functions that we consider canonical?

Starting with Blackwell monotonicity, take  $\pi, \rho \in \Pi(\mu)$  such that  $\pi \succeq \rho$ . Then, by setting  $f := -K$ , we obtain

$$\begin{aligned} -\langle K, \pi \rangle \geq -\langle K, \rho \rangle &\Rightarrow K(\mu) - \langle K, \pi \rangle \geq K(\mu) - \langle K, \rho \rangle \\ &\Rightarrow C(\pi) \geq C(\rho), \end{aligned}$$

with the second implication following from posterior separability. Hence, (C<sub>4</sub>) holds as desired. Switching now to convexity, take  $\pi, \rho \in \Pi(\mu)$  and  $\lambda \in (0, 1)$ . Then, we obtain

$$\begin{aligned} C(\lambda\pi + (1 - \lambda)\rho) &= K(\mu) - \langle K, \lambda\pi + (1 - \lambda)\rho \rangle \\ &= \lambda K(\mu) - \lambda \langle K, \pi \rangle + (1 - \lambda)K(\mu) - (1 - \lambda) \langle K, \rho \rangle \\ &= \lambda C(\pi) + (1 - \lambda)C(\rho), \end{aligned}$$

with the first and third equations following from posterior separability, and the second one following from the linearity of the inner product. Hence, our axioms imply an even stronger axiom than (C<sub>5</sub>), viz.,  $C$  is linear in  $\Pi(\mu)$ .

**Multinomial beliefs.** Throughout the paper we have focused on binary state spaces, thus eliciting the probability of a single event. The technical difficulty with directly extending our main result to the multinomial case lies on the extension of Lemma 2 to higher-dimension euclidean spaces not being straightforward. Nevertheless, for practical purposes, this problem is of minor concern. Indeed, on the one hand, whenever the function  $K$  is subdifferentiable at the boundary of  $\Delta(\Omega)$ , this lemma is not necessary and our Theorem 1 holds verbatim for any finite  $\Omega$ . Furthermore, our analysis on approximate robustness can be extended to the case of multinomial beliefs. Hence, our assumption on  $\Omega$  being binary is essentially without loss of generality.

**Violations of posterior-separability.** As we have already discussed earlier in the paper, posterior-separability has both solid axiomatic foundations and is supported by recent experimental evidence. Nevertheless, from a theoretical point of view, it is of some interest to understand whether our theory can be extended to broader classes of cost functions. It is quite obvious that violations of  $(C_2)$  always lead to non-existence of robust scoring rules. Nevertheless, depending on the particular violations of  $(C_2)$ , approximate robustness can be sometimes established, e.g., if  $C$  is characterized by a weakly concave  $K$ , then there exists some  $\delta > 0$  such that for every  $\varepsilon \geq \delta$  there is an  $\varepsilon$ -robust scoring rule. On the other hand, violations of  $(C_3)$  are more challenging to tackle, and would require a more systematic study in their own right. The technical reason is that in the absence of  $(C_3)$  we cannot necessarily characterize the costs with a function on the same domain as the benefits, and therefore completely different approaches have to be followed.

**Lying aversion.** Experimental evidence in different contexts suggests that, in the absence of explicit incentives, subjects often exhibit aversion to lying (Gneezy, 2005; Hurkens and Kartik, 2009). Whenever this is the case, every linear scoring rule would be robust, as long as his cost function satisfies  $(C_1) - (C_2)$ . For instance consider non-incentivized belief elicitation (i.e.,  $S_r(\omega) = 0$  for all  $\omega \in \Omega$  and  $r \in [0, 1]$ ), and observe that the subject will not update his beliefs as the expected benefits of paying attention will be equal to 0, and they will be therefore offset by the strictly positive costs that any attention strategy will induce. Then, due to lying aversion the subject will report his true belief, given that he will be indifferent between lying and telling the truth. Of course, it is important to stress that, specifically in the context of belief elicitation, the experimental evidence on the effects of monetary incentives are not conclusive and therefore it remains unclear whether we can rely on lying aversion.

**Testing our theory.** As we have already mentioned throughout the paper, there is no consensus in the literature as to whether monetary incentives are needed in order to elicit the subject's true beliefs. Furthermore, if incentives are needed, it is also unclear whether the strength of said incentives plays a crucial role, as our theory suggests. In either case, this is a question of systematic experimental testing. The most obvious way to proceed would be to vary the convexity of the scoring rule across treatments.

## A. Proofs of Section 3

### A.1. Intermediate results

**Proof of Lemma 1.** SUFFICIENCY. Let  $C$  be posterior separable. Then, it is obvious that  $C(\hat{\mu}) = 0$  for every  $\mu \in [0, 1]$ , thus proving  $(C_1)$ . By strict concavity of  $K$  it follows that  $C(\pi) = K(\mu) - \langle K, \pi \rangle > 0$  for every  $\pi \in \Pi(\mu)$  and every  $\mu \in [0, 1]$ , thus proving  $(C_2)$ . Finally, for an arbitrary

$\sigma : [0, 1] \rightarrow \Delta([0, 1])$  satisfying  $\sigma(\mu) \in \Pi(\mu)$  for every  $\mu \in [0, 1]$ , and an arbitrary  $\pi \in \Pi(\mu_0)$ ,

$$\begin{aligned} C(\pi) + \mathbb{E}_\pi(C \circ \sigma) &= K(\mu_0) - \langle K, \pi \rangle + \mathbb{E}_\pi(K - \langle K, \sigma \rangle) \\ &= K(\mu_0) - \langle K, \pi \rangle + \langle K, \pi \rangle - \langle K, \mathbb{E}_\pi(\sigma) \rangle \\ &= C(\mathbb{E}_\pi(\sigma)), \end{aligned}$$

with the first and the third equation following from posterior-separability, and the second one following from the linearity of the expectation ( $\mathbb{E}_\pi$ ) and the inner product ( $\langle K, \cdot \rangle$ ). Hence,  $(C_3)$  is also proven.

NECESSITY. Assume that  $C$  satisfies  $(C_1) - (C_3)$ , and let  $K : [0, 1] \rightarrow \mathbb{R}_+$  be the cost of learning the state with certainty, i.e.,  $K(\mu) := C(\sigma(\mu))$  for each  $\mu \in [0, 1]$ , where  $\sigma(\mu) \in \Pi(\mu)$  induces with probability  $\mu$  the posterior that puts probability 1 to  $\omega_0$  and with probability  $1 - \mu$  the posterior that puts probability 1 to  $\omega_1$ . Now, for an arbitrary  $\pi \in \Pi(\mu_0)$ ,

$$\begin{aligned} C(\pi) &= C(\mathbb{E}_\pi(\sigma)) - \mathbb{E}_\pi(C \circ \sigma) \\ &= K(\mu_0) - \langle K, \pi \rangle, \end{aligned}$$

with the first equation following directly from rearranging  $(C_3)$ , and the second one following from the definition of  $K$ . Hence, it suffices to prove that  $K$  is strictly concave. Take arbitrary  $0 \leq \mu_1 < \mu_2 \leq 1$  and  $\theta \in (0, 1)$ , and let  $\pi_0 \in \hat{\Pi}(\theta\mu_1 + (1 - \theta)\mu_2)$  be the attention strategy that assigns probability  $\theta$  to  $\mu_1$  and probability  $1 - \theta$  to  $\mu_2$ . Then,

$$\begin{aligned} K(\theta\mu_1 + (1 - \theta)\mu_2) &= C(\pi_0) + \theta K(\mu_1) + (1 - \theta)K(\mu_2) \\ &> \theta K(\mu_1) + (1 - \theta)K(\mu_2), \end{aligned}$$

with the equation above following from  $(C_3)$ , and the inequality following from  $(C_2)$ . Hence,  $K$  is strictly concave, thus completing the proof.  $\square$

**Proof of Lemma 2.** If  $f$  is subdifferentiable in  $[0, 1]$  then the result follows trivially by setting  $g := f$ . Therefore, we assume that there exists  $x \in \{0, 1\}$  such that the subderivative

$$\partial f(x) := \{t \in \mathbb{R} : f(y) \geq f(x) + t(y - x) \text{ for all } y \in [0, 1]\}$$

is empty.

STEP 1: By convexity,  $f$  is continuous in  $(0, 1)$ . Let  $\hat{f} : [0, 1] \rightarrow \mathbb{R}$  be the continuous extension of  $f : (0, 1) \rightarrow \mathbb{R}$  to  $[0, 1]$ . It is straightforward that  $\hat{f}$  exists and is strictly convex. Let us now prove that  $f - \hat{f}$  is convex. Take arbitrary  $0 \leq x_1 < x_2 \leq 1$  and  $\theta \in (0, 1)$ . Since  $(\theta x_1 + (1 - \theta)x_2) \in (0, 1)$ , we trivially obtain  $(f - \hat{f})(\theta x_1 + (1 - \theta)x_2) = 0$ . Moreover, by  $f(x) \geq \hat{f}(x)$ , we obtain  $\theta(f - \hat{f})(x_1) + (1 - \theta)(f - \hat{f})(x_2) \geq 0$ . Hence,  $f - \hat{f}$  is convex, as claimed. Therefore, it suffices to prove that there is a strictly convex and subdifferentiable  $g$  such that  $\hat{f} - g$  is convex.

STEP 2: For each  $x \in [0, 1]$  define the left  $a_x := \hat{f}'_-(x)$  and right  $b_x := \hat{f}'_+(x)$  derivative respectively. We adopt the notational convention that  $a_0 = -\infty$  and  $b_1 = \infty$ . It follows from (strict) convexity of  $\hat{f}$  that  $\partial \hat{f}(x) = [a_x, b_x]$ , with  $a_x = b_x$  whenever  $\hat{f}$  is differentiable at  $x$ . Moreover,  $\partial \hat{f}$  is strictly increasing, i.e.,  $x < y$  if and only if  $a_x \leq b_x < a_y \leq b_y$ . Obviously,  $\hat{f}$  is subdifferentiable if and only if  $-\infty < b_0 < a_1 < \infty$ , in which case we simply set  $g := \hat{f}$ . Hence, we henceforth focus on the case where  $\partial \hat{f}(x) = \emptyset$  for some  $x \in \{0, 1\}$ , i.e.,  $b_0 = -\infty$  or  $a_1 = \infty$ . Let  $x_0 \in [0, 1]$  be the unique minimizer of  $\hat{f}$ , and define the strictly increasing function  $F : [0, 1] \rightarrow \overline{\mathbb{R}}$  as follows:  $F(x) := a_x > 0$  for all  $x \in (x_0, 1]$ ,  $F(x) := b_x < 0$  for all  $x \in [0, x_0)$ , and  $F(x_0) = 0$ .

STEP 3: Since  $\hat{f}$  is continuous in a closed interval, it is also absolutely continuous, and therefore by the Fundamental Theorem of Calculus,  $F$  is Lebesgue integrable and

$$\hat{f}(x) = \hat{f}(0) + \int_0^x F(t)dt. \quad (\text{A.1})$$

Take a strictly increasing Lipschitz function  $h : \overline{\mathbb{R}} \rightarrow [-1, 1]$  (with Lipschitz constant  $c \leq 1$ ), and let  $G := h \circ F$ . Since  $F$  is Lebesgue integrable, so is  $G$ . Thus, we can define  $g : [0, 1] \rightarrow \mathbb{R}$  by

$$g(x) := \hat{f}(0) + \int_0^x G(t)dt. \quad (\text{A.2})$$

Since  $G$  is strictly increasing,  $g$  is strictly convex, and therefore subdifferentiable in  $(0, 1)$ . Moreover, since  $G$  takes values in  $[-1, 1]$ , it is the case that  $\int_0^x G(t)dt \geq -2x$ , implying that  $g(x) \geq g(0) - 2x$  for every  $x \in [0, 1]$ , i.e.,  $g$  is subdifferentiable at 0. We prove identically that  $g$  subdifferentiable also at 1, implying that it is subdifferentiable in  $[0, 1]$ .

STEP 4: Let us finally prove that  $\hat{f} - g$  is convex. Consider arbitrary  $0 \leq x_1 < x_2 \leq 1$ . Since  $h$  is Lipschitz with constant  $c \leq 1$ , it is the case that  $F(x_2) - F(x_1) \geq G(x_2) - G(x_1)$ , implying that  $F - G$  is increasing. Moreover, by (A.1) and (A.2), we obtain  $(\hat{f} - g)(x) = \int_0^x (F(t) - G(t))dt$ , implying that  $\hat{f} - g$  is convex, which completes the proof.  $\square$

## A.2. Proof of Theorem 1: Main existence result

First, observe that a proper scoring rule  $\phi$  is robust if and only if  $\phi + K$  is strictly concave. Indeed, for arbitrary  $\pi \in \hat{\Pi}(\mu)$  and  $\mu \in [0, 1]$ ,

$$\begin{aligned} V_\phi(\pi) &= \langle \phi, \pi \rangle - \phi(\mu) - K(\mu) + \langle K, \pi \rangle \\ &= \langle \phi + K, \pi \rangle - (\phi + K)(\mu), \end{aligned}$$

with the first equation following from Lemma 1 and the definition of  $V_\phi$ . Now for arbitrary  $\gamma \in \mathbb{R}$  and  $\lambda \in (0, 1)$ , define the strictly convex function  $f := \gamma - \lambda K$  (see Equation (6)). Then, by Lemma 2, there exists some strictly convex and subdifferentiable function  $g$ , such that  $f - g$  is convex. Set  $\phi := g$ . By strict convexity and subdifferentiability of  $g$ , it follows that  $\phi$  is a proper scoring rule. Finally, notice that  $g + K$  is strictly concave, as it is the sum of a strictly concave function (viz.,  $K + f$ ) and a concave function (viz.,  $g - f$ ). Therefore, by our first argument,  $\phi$  is robust.

## A.3. Proof of Proposition 1: Entropic costs

Recall from the proof of Theorem 1 that  $\phi_\beta$  is robust if and only if  $\psi_\beta := \phi_\beta + \kappa H$  is strictly concave. Observe that  $\psi_\beta''(\mu) = 2\beta - \frac{\kappa}{\mu(1-\mu)}$ , implying that  $\psi_\beta''(\mu) \leq 0$  if and only if  $2\beta\mu^2 - 2\beta\mu + \kappa \geq 0$ . Take  $\Delta := 4\beta^2 - 8\beta\kappa$  and observe that there are three possible cases. If  $\Delta < 0$  then  $v''(\mu) < 0$  for all  $\mu \in [0, 1]$ . If  $\Delta = 0$  then  $\psi_\beta''(\mu) \leq 0$  for all  $\mu \in [0, 1]$  with equality holding only for  $\mu = 1/2$ . Finally, if  $\Delta > 0$  then  $v''(\mu) > 0$  for all  $\mu \in [0, 1] \cap (\frac{1}{2} - \frac{\sqrt{\Delta}}{4\beta}, \frac{1}{2} + \frac{\sqrt{\Delta}}{4\beta})$ . Hence,  $v$  is strictly concave in  $[0, 1]$  if and only if  $\Delta \leq 0$ , which is the case if and only if  $\beta \leq 2\kappa$ .

## B. Proofs of Section 4

### B.1. Proof of Proposition 2: Quadratic scoring rules

Since  $K$  is a strictly concave polynomial of degree  $n$ , the (negative) second derivative has  $k \leq n - 2$  roots. Obviously, if  $K''$  has no root then by continuity,  $\max\{K''(\mu) | \mu \in [0, 1]\} < 0$ , implying that

there exists some  $\beta > 0$  such that  $K + \phi_\beta$  is strictly concave, and therefore  $\phi_\beta$  is robust. So let focus on the case where  $K''$  has at least one root. Fix an arbitrary  $\varepsilon > 0$ .

STEP 1: Since  $K'' \leq 0$ , the roots of  $K''$  are also its maxima, viz.,  $\arg \max_{\mu \in [0,1]} K''(\mu) = \{\mu_1, \dots, \mu_k\}$ . Hence, there exists some  $\hat{\delta} > 0$  such that  $K''$  is strictly increasing in  $[\hat{\mu}_j^1, \mu_j]$  and strictly decreasing in  $[\mu_j, \hat{\mu}_j^2]$  for every  $\mu_j$ , where  $\hat{\mu}_j^1 := \max\{\mu_j - \hat{\delta}, 0\}$  and  $\hat{\mu}_j^2 := \min\{\mu_j + \hat{\delta}, 1\}$ . Then, define  $\hat{\beta} := \max\{K''(\hat{\mu}_j^\ell) | j \in \{1, \dots, k\} \text{ and } \ell \in \{1, 2\}\}$ . Then, by continuity, for each  $\beta \in (0, \hat{\beta})$  and each  $\mu_j$ , there are unique  $\mu_j^{\beta,1} \in (\hat{\mu}_j^1, \mu_j)$  and  $\mu_j^{\beta,2} \in (\mu_j, \hat{\mu}_j^2)$  such that  $K''(\mu_j^{\beta,1}) + \beta = K''(\mu_j^{\beta,2}) + \beta = 0$ . Henceforth denote the corresponding interval by  $I_j^\beta := [\mu_j^{\beta,1}, \mu_j^{\beta,2}]$ , and observe that  $K''(\mu) + \beta \geq 0$  for all  $\mu \in I_j^\beta$ , with strict inequality holding if and only if  $\mu \in \text{int}(I_j^\beta)$ .

STEP 2: Obviously  $\mu_j^{\beta,\ell}$  converges continuously to 0 as  $\beta \rightarrow 0$ . Hence, there exists some  $\beta^* > 0$  such that for all  $\beta \in (0, \beta^*)$  it is the case that  $K''(\mu) < \beta$  for all  $\mu \notin I^\beta := I_1^\beta \cup \dots \cup I_k^\beta$ . Hence for all  $\beta \in (0, \beta^*)$ , the function  $\psi_\beta := K + \phi_\beta$  is strictly convex in every  $I_j^\beta$  and strictly concave in every convex subset of  $\text{clos}([0, 1] \setminus I^\beta)$ . Now take the concave closure  $\bar{\psi}_\beta(\mu) := \sup\{\langle \psi_\beta, \pi \rangle | \pi \in \Pi(\mu)\}$  of  $\psi_\beta$  (Kamenica and Gentzkow, 2011). Notice that by construction  $\bar{\psi}_\beta$  is linear in a closed interval  $J_j^\beta := [\nu_j^{\beta,1}, \nu_j^{\beta,2}] \supseteq I_j^\beta$  for each  $I_j^\beta$ , and strictly concave in every convex subset of  $\text{clos}([0, 1] \setminus I^\beta)$ .

STEP 3: Now for every  $j \in \{1, \dots, k\}$  and every  $\ell \in \{1, 2\}$ , we prove that  $\nu_j^{\beta,\ell}$  converges monotonically to  $\mu_j$  as  $\beta \rightarrow 0$ . Without loss of generality, we prove the previous claim for  $\ell = 1$ . Suppose that this is not the case. Then, there exists some  $\nu_j^1 \in (\hat{\mu}_j^1, \mu_j)$  such that  $\nu_j^{\beta,1} < \nu_j^1$  for all  $\beta \in (0, \beta^*)$ . Then, for every  $\lambda \in (0, 1)$  it is the case that  $\bar{\psi}_\beta(\lambda \nu_j^1 + (1 - \lambda)\mu_j) = \lambda \bar{\psi}_\beta(\nu_j^1) + (1 - \lambda)\bar{\psi}_\beta(\mu_j)$ . Moreover, observe that  $\bar{\psi}_\beta(\mu) \rightarrow K(\mu) + \alpha$  as  $\beta \rightarrow 0$  for every  $\mu \in [0, 1]$ , implying that

$$\begin{aligned} K(\lambda \nu_j^1 + (1 - \lambda)\mu_j) &= \lim_{\beta \rightarrow 0} \bar{\psi}_\beta(\lambda \nu_j^1 + (1 - \lambda)\mu_j) - \alpha \\ &= \lim_{\beta \rightarrow 0} (\lambda \bar{\psi}_\beta(\nu_j^1) + (1 - \lambda)\bar{\psi}_\beta(\mu_j)) - \alpha \\ &= \lambda K(\nu_j^1) + (1 - \lambda)K(\mu_j), \end{aligned}$$

which contradicts the fact that  $K$  is strictly concave.

STEP 4: By construction, for every  $\beta > 0$ , if  $\mu \in J_j^\beta$  then  $\arg \max_{\pi \in \Pi(\mu)} V_{\phi_\beta}(\pi) = \{\pi_j\}$  where  $\text{supp}(\pi_j) = \{\nu_j^1, \nu_j^2\}$ . On other hand, if  $\mu \notin J^\beta := J_1^\beta \cup \dots \cup J_k^\beta$  then  $\arg \max_{\pi \in \Pi(\mu)} V_{\phi_\beta}(\pi) = \{\hat{\mu}\}$ . By the previous step, there exists some  $\beta \in (0, \beta^*)$  such that  $\nu_j^{\beta,2} - \nu_j^{\beta,1} \leq \varepsilon$  for all  $j \in \{1, \dots, k\}$ . Hence,  $\nu \in B_\varepsilon(\mu)$ , for all  $\nu \in \text{supp}(\pi)$  with  $\pi \in \arg \max_{\rho \in \Pi(\mu)} V_{\phi_\beta}(\rho)$ . Therefore,  $\phi_\beta$  is  $\varepsilon$ -robust.

## B.2. Proof of Proposition 3: Discrete scoring rules

Take the collection  $(\phi_k)_{k=0}^n$  of linear functions  $\phi_k(\mu) := c_k + d_k \mu$ , and define the discrete scoring rule  $S : R \times \Omega \rightarrow \mathbb{R}$  by  $S_{r_k}(\omega_0) := \phi_k(1)$  and  $S_{r_k}(\omega_1) := \phi_k(0)$  for each  $r_k \in R$ . Let us first construct the  $\phi_k$ 's in a way such that  $S$  is a  $(1/n)$ -robust discrete scoring rule. For each  $k \in \{1, \dots, n\}$  define

$$\mu_k := \frac{2k - 1}{2n}, \tag{B.1}$$

subsequently observing that  $\mu_k = (r_{k-1} + r_k)/2$ . Then, for every  $\mu_k$ , consider arbitrary  $\mu_k^1 \in (r_{k-1}, \mu_k)$  and  $\mu_k^2 \in (\mu_k, r_k)$  such that  $\mu_k^2 - \mu_k^1 = \delta < 1/2n$  for every  $k \in \{1, \dots, n\}$ . By strict concavity of

$K$ , we obtain  $K'_-(\mu_k^1) > K'_+(\mu_k^2) > K'_-(\mu_{k+1}^1) > K'_+(\mu_{k+1}^2)$ . Then, we inductively define  $(c_k, d_k)$  as follows: Fix arbitrary  $c_0 \in \mathbb{R}$  and  $d_0 \in \mathbb{R}$ , and given  $(c_k, d_k)$  define

$$d_{k+1} := d_k + K'_+(\mu_{k+1}^2) - K'_-(\mu_{k+1}^1), \quad (\text{B.2})$$

$$c_{k+1} := c_k + (d_k - d_{k+1})\mu_{k+1}, \quad (\text{B.3})$$

for each  $k \in \{0, \dots, n-1\}$ . Note that by construction,  $\phi_{k-1}(\mu_k) = \phi_k(\mu_k)$  for every  $k \in \{1, \dots, n\}$ .

**APPROXIMATE PROPERNESS.** Let us prove that  $\phi := \max\{\phi_1, \dots, \phi_n\}$  is  $(1/2n)$ -proper. Notice that  $\phi_k(\mu) \geq \phi_\ell(\mu)$  for all  $\ell \in \{0, \dots, n\}$  if and only if  $\mu \in [r_{k-1}, r_k]$ , with equality holding if and only if  $\mu \in \{r_{k-1}, r_k\}$ . The latter follows from the fact that (i)  $d_k$  is strictly increasing in  $k$ , and (ii)  $\phi_{k-1}(\mu_k) = \phi_k(\mu_k)$  for all  $k \in \{1, \dots, n\}$ . Hence,  $\phi$  is piecewise linear and convex with a kink at every  $\mu \in \{\mu_1, \dots, \mu_n\}$ . Hence,  $\mathbb{E}_\mu(S_{r_k}) > \mathbb{E}_\mu(S_r)$  for every  $\mu \in (\mu_k, \mu_{k+1})$  and every  $r \in R \setminus \{r_k\}$ . Moreover,  $\mathbb{E}_{\mu_k}(S_{r_{k-1}}) = \mathbb{E}_{\mu_k}(S_{r_k}) > \mathbb{E}_{\mu_k}(S_r)$  and every  $r \in R \setminus \{r_{k-1}, r_k\}$ . Therefore,  $\phi$  is  $(1/2n)$ -proper.

**APPROXIMATE ROBUSTNESS.** Take the function  $\psi := K + \phi$ , and consider its concave closure  $\bar{\psi}(\mu) := \sup\{\langle \psi, \pi \rangle \mid \pi \in \Pi(\mu)\}$  (Kamenica and Gentzkow, 2011). By construction, for every  $k \in \{1, \dots, n\}$ ,

$$\begin{aligned} \bar{\psi}'_+(\mu_k^1) &= K'_+(\mu_k^1) + d_{k-1} \\ &= K'_-(\mu_k^2) + d_k \\ &= \bar{\psi}'_-(\mu_k^2), \end{aligned}$$

implying that  $\bar{\psi}$  is linear in each interval  $[\mu_k^1, \mu_k^2]$ , and it is strictly concave in every convex subset of  $[0, 1] \setminus M$ , where  $M := (\mu_1^1, \mu_1^2) \cup \dots \cup (\mu_n^1, \mu_n^2)$ . Hence, we consider two cases:

- (i)  $\mu \in (\mu_k^1, \mu_k^2)$ : the subject's optimal attention strategy is  $\pi_k \in \Pi(\mu)$  such that  $\text{supp}(\pi_k) = \{\mu_k^1, \mu_k^2\}$ . Hence, by the previous step, upon forming the posterior  $\mu_k^1$  the subject will report  $r_{k-1}$ , whereas upon forming the posterior  $\mu_k^2$  he will report  $r_k$ .
- (ii)  $\mu \in [0, 1] \setminus M$ : the subject's optimal attention strategy is  $\hat{\mu}$ . Since, by construction  $\mu \in [\mu_k^2, \mu_{k+1}^1]$ , again by the previous step, the subject will report  $r_k$ .

Therefore, we conclude that  $\phi$  is  $(1/n)$ -robust, thus completing the proof.

## C. Proofs of Section 5

### C.1. Intermediate results

We begin by showing that our  $(\varepsilon, \delta)$ -robustness is well-defined. In particular, the following result guarantees that for every convex and continuous  $\phi : [0, 1] \rightarrow \mathbb{R}$  and every  $\varepsilon \geq 0$ , the event  $\{K \in \mathcal{C} : \varepsilon_\phi^K \leq \varepsilon\}$  is measurable in the Borel  $\sigma$ -algebra generated by the topology of the sup norm.

**Lemma C1.** *For every weakly proper scoring rule  $\phi$ , the function  $\varepsilon_\phi$  is upper semi-continuous.*

**PROOF.** Fix an arbitrary scoring rule  $\phi$ , and take an arbitrary sequence  $(K_t)_{t=1}^\infty$  in  $\mathcal{K}$  converging to some  $K_0 \in \mathcal{K}$  in the topology induced by the sup norm.

**STEP 1:** By  $K_t - K_0 = \psi_t - \psi_0$ , we obtain  $\psi_t \rightarrow \psi_0$ , where as usual  $\psi_t := \phi + K_t$ . That is, for every  $\delta > 0$  there exists some  $t_\delta \in \mathbb{N}$  such that  $\|\psi_t - \psi_0\|_\infty := \sup_{\mu \in [0, 1]} |\psi_t(\mu) - \psi_0(\mu)| < \delta$  for every  $t > t_\delta$ .

Now, for every  $t \in \mathbb{N}$ , take the concave closure  $\bar{\psi}_t(\mu) := \sup_{\pi \in \Pi(\mu)} \langle \psi_t, \pi \rangle$ , and observe that

$$\begin{aligned} |\bar{\psi}_t(\mu) - \bar{\psi}_0(\mu)| &= \left| \sup_{\pi \in \Pi(\mu)} \langle \psi_t, \pi \rangle - \sup_{\pi \in \Pi(\mu)} \langle \psi_0, \pi \rangle \right| \\ &\leq \left| \sup_{\pi \in \Pi(\mu)} \langle \psi_t - \psi_0, \pi \rangle \right| \\ &\leq \sup_{\pi \in \Pi(\mu)} |\langle \psi_t - \psi_0, \pi \rangle| \\ &\leq \sup_{\pi \in \Pi(\mu)} \langle |\psi_t - \psi_0|, \pi \rangle. \end{aligned}$$

Then, by the definition of the sup norm, we obtain

$$\begin{aligned} \|\bar{\psi}_t - \bar{\psi}_0\|_\infty &\leq \sup_{\mu \in [0,1]} \sup_{\pi \in \Pi(\mu)} \langle |\psi_t - \psi_0|, \pi \rangle \\ &\leq \sup_{\mu \in [0,1]} |\psi_t - \psi_0| \\ &= \|\psi_t - \psi_0\|_\infty, \end{aligned}$$

implying that  $\bar{\psi}_t \rightarrow \bar{\psi}_0$ .

**STEP 2:** Fix an arbitrary  $\mu \in (0, 1)$ . For each  $t \in \mathbb{N}$ , define the largest closed interval  $[a_t, b_t]$  containing  $\mu$ , such that  $\bar{\psi}_t$  is linear in  $[a_t, b_t]$ . Moreover, for each  $\nu \in [0, 1]$  define the largest closed interval  $I_\phi(\nu) := [a_\phi(\nu), b_\phi(\nu)]$  containing  $\nu$  such that  $\phi$  is linear in  $I_\phi(\nu)$ . Obviously, when the subject's posterior belief is  $\nu$ , every report in  $[a_\phi(\nu), b_\phi(\nu)]$  yields the same expected utility  $\phi(\nu)$ . Notice that both  $a_\phi$  and  $b_\phi$  are decreasing in  $\nu$ . Then, we define  $a_t^* := a_\phi(a_t)$  and  $b_t^* := b_\phi(b_t)$ , i.e., the interval  $[a_t^*, b_t^*]$  is the smallest closed interval that contains all the posteriors that can be optimally reported by the subject when his prior is  $\mu$ . In other words,  $a_t^*$  and  $b_t^*$  are the worst-case scenarios, in terms of distance from  $\mu$ .

We will now show that

$$a_0^* \leq \liminf a_t^* \leq \limsup b_t^* \leq b_0^*. \quad (\text{C.1})$$

For every  $t \in \mathbb{N}$ , the (most dispersed) optimal attention strategy at  $\mu$  is denoted by  $\pi_t \in \Pi(\mu)$  and is distributed over  $\{a_t, b_t\}$ . Hence, we obtain  $\bar{\psi}_t(\mu) = \langle \psi_t, \pi_t \rangle$ ,<sup>10</sup> and therefore, by Step 1,

$$\langle \psi_t, \pi_t \rangle \rightarrow \langle \psi_0, \pi_0 \rangle. \quad (\text{C.2})$$

Now suppose – contrary to what we want to show – that  $\liminf a_t^* < a_0^*$  or  $\limsup b_t^* > b_0^*$ . Since  $\inf_{k \geq t} a_k^*$  is increasing and  $\sup_{k \geq t} b_k^*$  is decreasing in  $k$ , there exists a subsequence of  $(a_t^*, b_t^*)$ , identified by a countable subset  $T \subseteq \mathbb{N}$ , such that for each  $t \in T$  it is the case that  $a_t^* < a_0^* - \delta^*$  or  $b_t^* > b_0^* + \delta^*$ , for some  $\delta^* > 0$ . The latter implies that  $I_\phi(a_t) \cap I_\phi(a_0) = \emptyset$  and  $I_\phi(b_t) \cap I_\phi(b_0) = \emptyset$  for all  $t \in T$ . Let us now consider two cases:

- (i)  $a_0 > a_0^*$ , i.e.,  $a_0$  is in the interior or at the upper bound of  $I_\phi(a_0)$ : Then, there is some  $\delta_a > 0$  such that  $a_t < a_0 - \delta_a$ . Likewise, if  $b_0 < b_0^*$ , there exists  $\delta_b > 0$  such that  $b_t > b_0 + \delta_b$ .

<sup>10</sup>Here we implicitly assume that  $\langle \psi_t, \rho \rangle$  achieves a maximum in  $\Pi(\rho)$ , which in turn is true when  $\psi_t$  is continuous. Assuming the latter is without loss of generality, because we can simply focus on continuous  $K$ 's. Indeed, discontinuities of  $K$  can only occur at the boundary of  $[0, 1]$ , yielding a lower semi-continuous  $K$ . Still, such a cost would dominate the continuous extension of  $K$  from the interior  $(0, 1)$  to the entire unit interval  $[0, 1]$ . Hence, if a scoring rule is robust (resp., approximately robust) under the continuous cost function, it will also be robust (resp., approximately robust) under the original discontinuous cost function.



- (ii)  $a_0 = a_0^*$ , i.e.,  $a_0$  is at the lower bound of  $I_\phi(a_0)$ : Then, there exists a neighborhood  $B_{\delta_a}(a_0)$  such that  $\phi$  is strictly convex in the left part of the neighborhood, viz., in  $\{\nu \in B_{\delta_a}(a_0) : \nu \leq a_0\}$ ; otherwise,  $a_0^* < a_0$  and we are back to the previous case. Hence,  $a_t < a_0 - \delta_a$ . Likewise, if  $b_0 = b_0^*$ , there is some  $\delta_b > 0$  such that  $b_t > b_0 + \delta_b$ .

In either case, there exists some strictly positive  $\delta < \min\{\delta_a, \delta_b\}$  such that  $a_t < a_0 - \delta$  or  $b_t > b_0 + \delta$ . Now, define the attention strategy  $\pi \in \Pi(\mu)$  with  $\text{supp}(\pi) = \{a_0 - \delta, b_0 + \delta\}$ , and observe that  $\langle \psi_0, \pi \rangle < \langle \psi_0, \pi_0 \rangle$ , while  $\langle \psi_t, \pi \rangle = \langle \psi_t, \pi_t \rangle$  for all  $t \in T$ . Since,  $\langle \psi_t, \pi \rangle \rightarrow \langle \psi_0, \pi \rangle$ , the latter obviously contradicts (C.2), thus proving (C.1).

STEP 3: Define first  $d_t^*(\mu) := b_t^* - a_t^*$ , and subsequently  $\varepsilon_t := \sup_{\mu \in [0,1]} d_t^*(\mu)$ . Then, it is straightforward to verify that  $\varepsilon_t = \varepsilon_\phi^{Kt}$ . Then, by Step 2,

$$\begin{aligned} d_0^*(\mu) &= b_0^* - a_0^* \\ &\geq \limsup b_t^* - \liminf a_t^* \\ &\geq \limsup (b_t^* - a_t^*) \\ &= \limsup d_t^*(\mu). \end{aligned}$$

Therefore, it follows directly that  $\limsup \varepsilon_t \leq \varepsilon_0$ , which completes the proof.  $\square$

## C.2. Proof of Theorem 2: Cost uncertainty

STEP 1: Fix an arbitrary  $\varepsilon > 0$  and take some  $n \in \mathbb{N}$  such that  $1/n \leq \varepsilon/2$ . Similarly to our earlier proof (of Proposition 3), we define  $r_k := k/n$  for each  $k \in \{0, \dots, n\}$ .

For an arbitrary  $\gamma > 0$ , define the linear function  $\phi_k^\gamma(\mu) := c_k^\gamma + d_k^\gamma \mu$  for each  $k \in \{1, \dots, n\}$ , where  $c_k^\gamma := -\frac{k(k-1)}{2n} \gamma$  and  $d_k^\gamma := k\gamma$ . Then, consider the piecewise linear function  $\phi^\gamma := \max\{\phi_1^\gamma, \dots, \phi_n^\gamma\}$ , observing that it is (weakly) convex, and it has a kink at every  $\mu \in \{r_1, \dots, r_{n-1}\}$ . The latter holds, because  $\phi_k^\gamma(r_k) = \phi_{k+1}^\gamma(r_k)$ .

Fix an arbitrary  $K \in \mathcal{K}$ , and define  $\mu_k := (r_k + r_{k-1})/2$  for each  $k \in \{1, \dots, n\}$ . Let  $L_k(\mu) := a_k + b_k \mu$  be an arbitrary tangent of  $K$  at  $\mu_k$ . By strict concavity of  $K$ , we obtain  $L_k(\mu) \geq K(\mu)$ , with equality holding only at  $\mu = \mu_k$ .

STEP 2: We will prove that there exists some  $\gamma_K > 0$  such that  $L_k(\mu) + \phi_k^\gamma(\mu) \geq \psi^\gamma(\mu)$ , for all  $k \in \{1, \dots, n\}$ , for all  $\gamma < \gamma_K$ , and for all  $\mu \in [0, 1]$  with equality holding only at  $\mu = \mu_k$ , where as usual  $\psi^\gamma := \phi^\gamma + K$ . Let  $\mu \in [r_{\ell-1}, r_\ell]$ . If  $\ell = k$ , then  $\psi^\gamma(\mu) = \phi_k^\gamma(\mu) + K(\mu)$ , and our claim follows directly from  $L_k$  being a tangent of  $K$  at  $\mu_k$ . Hence, we let  $\ell \neq k$ , letting without loss of generality  $\ell > k$  (the proof is analogous for  $\ell < k$ ). Hence, it is the case that  $\mu \geq r_k > \mu_k$ , and therefore

$$L_k(\mu) - K(\mu) \geq L_k(r_k) - K(r_k) > 0,$$

noticing that  $L_k(r_k) - K(r_k)$  does not depend either on  $\gamma$  or on  $\mu$ . Moreover, by construction,

$$\begin{aligned} 0 &\leq \phi_\ell^\gamma(\mu) - \phi_k^\gamma(\mu) \\ &< \phi_\ell^\gamma(1) - \phi_k^\gamma(1) \\ &\leq D_k \gamma, \end{aligned}$$

where  $D_k := \max\{(\ell - k)(2n + 1 - \ell - k) \mid \ell \in \{k + 1, \dots, n\}\} / 2n > 0$ . Note that  $D_k$  does not depend on  $\mu$ . Hence, we have

$$\begin{aligned} L_k(\mu) + \phi_k^\gamma(\mu) - \psi(\mu) &= L_k(\mu) + \phi_k^\gamma(\mu) - K(\mu) - \phi_\ell^\gamma(\mu) \\ &> (L_k(r_k) - K(r_k)) - D_k \gamma. \end{aligned}$$

Hence, by continuity, there exists some  $\gamma_K > 0$  such that  $(L_k(r_k) - K(r_k)) - D_k\gamma > 0$  for all  $\gamma < \gamma_K$ , as claimed above.

STEP 3: Take  $\gamma < \gamma_K$ . Then, by Step 2, we obtain  $\bar{\psi}^\gamma(\mu_k) = \psi^\gamma(\mu_k)$ , implying that the optimal attention strategy at  $\mu_k$  is  $\hat{\mu}_k$ . Now consider some  $\mu \in (\mu_k, \mu_{k+1})$ , and let  $\pi \in \Pi(\mu)$  be the optimal attention strategy with the most dispersed posteriors, i.e.,  $\text{supp}(\pi) = \{\nu_1, \nu_2\}$  where  $[\nu_1, \nu_2]$  is the largest interval of  $\mu$  where  $\bar{\psi}^\gamma$  is linear. Hence,  $\lambda\psi^\gamma(\nu_1) + (1 - \lambda)\psi^\gamma(\nu_2) \geq \psi^\gamma(\lambda\nu_1 + (1 - \lambda)\nu_2)$  for every  $\lambda \in (0, 1)$ . Therefore, in order not to contradict Step 2, it must necessarily be the case that  $[\nu_1, \nu_2] \subseteq [\mu_k, \mu_{k+1}]$ . Similarly, if  $\mu \in [0, \mu_1)$  every optimal attention strategy yields posteriors in  $[0, \mu_1]$ , and likewise if  $\mu \in (\mu_n, 1]$  every optimal attention strategy yields posteriors in  $[\mu_n, 1]$ . Finally, if the posterior is  $\mu = r_k$  then any report in report  $[r_{k-1}, r_{k+1}]$  is optimal, whereas if the posterior is  $\mu \in (r_k, r_{k+1})$  then every report in  $[r_k, r_{k+1}]$  is optimal. Therefore,  $\phi^\gamma$  is  $\varepsilon$ -robust given the cost  $K$ .

STEP 4: Take a strictly decreasing sequence  $(\gamma_t)_{t=1}^\infty$  such that  $\gamma_t \downarrow 0$ . Then, for every  $K \in \mathcal{K}$ , define  $T_K := \min\{t \in \mathbb{N} : \phi^{\gamma_t} \text{ is } \varepsilon\text{-robust given } K\}$ . Now, for each  $t \in \mathbb{N}$  define

$$\mathcal{K}_t := \{K \in \mathcal{K} : T_K \leq t\}, \quad (\text{C.3})$$

trivially noticing that  $\mathcal{K}_t \subseteq \mathcal{K}_{t+1}$ . Moreover, since  $T_K$  exists for every  $K \in \mathcal{K}$  (by Step 3), it is obviously the case that  $\mathcal{K}_t \uparrow \mathcal{K}$ . Therefore, by Billingsley (1995, Thm 2.1), we obtain  $P(\mathcal{K}_t) \uparrow p$ , implying that for every  $\delta > 1 - p$  there is some  $t_\delta \in \mathbb{N}$  such that  $P(\mathcal{K}_t) \geq 1 - \delta$  for all  $t \geq t_\delta$ .

STEP 5: For every  $t \in \mathbb{N}$ , define the event

$$\mathcal{E}_t := \{K \in \mathcal{C} : \varepsilon_{\phi^{\gamma_t}}^K \leq \varepsilon\}, \quad (\text{C.4})$$

noticing  $\mathcal{K}_t = \mathcal{E}_t \cap \mathcal{K}$ . Indeed, for every  $t \in \mathbb{N}$  and  $K \in \mathcal{K}$ ,

$$\begin{aligned} K \in \mathcal{K}_t &\Leftrightarrow \phi^{\gamma_t} \text{ is } \varepsilon\text{-robust given } K \\ &\Leftrightarrow \varepsilon_{\phi^{\gamma_t}}^K \leq \varepsilon \\ &\Leftrightarrow K \in \mathcal{E}_t, \end{aligned}$$

thus implying (by Step 4) that for every  $\delta > 1 - p$  there is some  $\phi \in \{\phi^{\gamma_1}, \phi^{\gamma_2}, \dots\}$  such that  $P(\{K \in \mathcal{C} : \varepsilon_\phi^K \leq \varepsilon\}) \geq 1 - \delta$ , thus completing the proof.

## References

- ALAOUI, L. & PENTA, A. (2018). Cost-benefit analysis in reasoning. *Working Paper*.
- ALLAIS, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica* 21, 503–546.
- ARMANTIER, O. & TREICH, N. (2013). Eliciting beliefs: proper scoring rules, incentives, stakes and hedging. *European Economic Review* 62, 17–40.
- BARTOŠ, V., BAUER, M., CHYTILOVÁ, J. & MATEJKA, F. (2016). Attention discrimination: theory and field experiments with monitoring information acquisition. *American Economic Review* 106, 1437–1475.
- BERGEMANN, D., MORRIS, S. & TAKAHASHI, S. (2017). Interdependent preferences and strategic distinguishability. *Journal of Economic Theory* 168, 329–371.

- BILLINGSLEY, P. (1995). *Probability and measure*. John Wiley & Sons.
- BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics* 24, 265–272.
- BRIER, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78, 1–3.
- CAPLIN, A. (2016). Measuring and modeling attention. *Annual Review of Economics* 8, 379–403.
- CAPLIN, A. & DEAN, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review* 105, 2183–2203.
- CAPLIN, A., DEAN, M. & LEAHY, J. (2017). Rationally inattentive behavior: characterizing and generalizing Shannon entropy. *Working Paper*.
- CHAMBERS, C. (2008). Proper scoring rules for general decision models. *Games and Economic Behavior* 63, 32–40.
- CHAMBERS, C. & LAMBERT, N. (2017). Dynamic belief elicitation. *Working Paper*.
- CHAMBERS, C., HEALY, P. & LAMBERT, N. (2017). Dual scoring. *Working Paper*.
- CHAMBERS, C., LIU, C. & REHBECK, J. (2017). Nonseparable costly attention and revealed preference. *Working Paper*.
- CLEMEN, R. (2002). Incentive contracts and strictly proper scoring rules. *Sociedad de Estadística e Investigación Operativa (Test)* 11, 167–189.
- COSTA-GOMES, M. & WEIZSÄCKER, G. (2008). Stated beliefs and play in normal-form games. *Review of Economic Studies* 75, 729–762.
- COVER, T.M. & THOMAS, J.A. (2006). *Elements of information theory*. John Wiley & Sons.
- DE OLIVEIRA, H., DENTI, T., MIHM, M. & OZBEK, K. (2017). Rationally inattentive preferences and hidden information costs. *Theoretical Economics* 12, 621–654.
- DEAN, M. & NELIGH, N. (2017). Experimental tests of rational inattention. *Working Paper*.
- ELLIS, A. (2018). Foundations for optimal inattention. *Journal of Economic Theory* 73, 56–94.
- GNEEZY, U. (2005). Deception: the role of consequences. *American Economic Review* 95, 384–394.
- GNEITING, T. & RAFTERY, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102, 359–378.
- GOOD, I.J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B*, 14, 107–114.
- GRISLEY, W. & KELLOGG, E.D. (1983). Farmers’ subjective probabilities in Northern Thailand: an elicitation analysis. *American Journal of Agricultural Economics* 65, 74–82.
- HANSON, R. (2003). Combinatorial information market design. *Information Systems Frontiers* 5, 107–119.

- 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.
- HÉBERT, B. & WOODFORD, M. (2016). Rational inattention with sequential information sampling. *Working Paper*.
- HOSSAIN, T. & OKUI, R. (2013). The binarized scoring rule. *Review of Economic Studies* 80, 984–1001.
- HURD, M. (2009). Subjective probabilities in household surveys. *Annual Review of Economics* 1, 543–562.
- HURKENS, S. & KARTIK, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics* 12, 180–192.
- KAMENICA, E. & GENTZKOW, M. (2011). Bayesian persuasion. *American Economic Review* 101, 2590–2615.
- KARNI, E. (1999). Elicitation of subjective probabilities when preferences are state-dependent. *International Economic Review* 40, 479–486.
- (2009). A mechanism for eliciting probabilities. *Econometrica* 77, 603–606.
- (2017). A mechanism for the elicitation of second-order belief and subjective information structure. *Working Paper*.
- MACCHERONI, F., MARINACCI, M. & RUSTICHINI, A. (2006). Dynamic variational preferences. *Journal of Economic Theory* 128, 4–44.
- MAĆKOWIAK, B. & WIEDERHOLT, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review* 99, 769–803.
- MANSKI, C.F. (2004). Measuring expectations. *Econometrica* 72, 1329–1376.
- MCKELVEY, R. & PAGE, T. (1990). Public and private information: an experimental study of information pooling. *Econometrica* 58, 1321–1339.
- MCCARTHY, J. (1956). Measures of the value of information. *Proceedings of the National Academy of Sciences* 42, 654–655.
- MATEJKA, F. (2016). Rationally inattentive seller: sales and discrete pricing. *Review of Economic Studies* 83, 1156–1188.
- MATEJKA, F. & MCKAY, A. (2015). Rational inattention to discrete choices: a new foundation for the multinomial logit model. *American Economic Review* 105, 272–298.
- MATEJKA, F. & TABELLINI, G. (2016). Electoral competition with rational inattentive voters. *Working Paper*.

- MORRIS, S. & STRACK, P. (2017). The Wald problem and the equivalence of sequential sampling and static information costs. *Working Paper*.
- NYARKO, Y. & SCHOTTER, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica* 70, 971–1005.
- 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.
- OSTROVSKY, M. (2012). Information aggregation in dynamic markets with strategic traders. *Econometrica* 80, 2595–2647.
- PALFREY, T. & WANG, S. (2009). On eliciting beliefs in strategic games. *Journal of Economic Behavior and Organization* 71, 98–109.
- RUTSTRÖM, E. & WILCOX, N. (2009). Stated beliefs versus inferred beliefs: a methodological inquiry and experimental test. *Games and Economic Behavior* 67, 616–632.
- SAVAGE, L. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* 66, 783–801.
- SCHLAG, K., TREMEWAN, J. & VAN DER WEELE, J. (2015). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics* 18, 457–490.
- SCHLAG, K. & VAN DER WEELE, J. (2013). Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality. *Theoretical Economics Letters* 3, 38–42.
- SCHOTTER, A. & TREVINO, I. (2014). Belief elicitation in the laboratory. *Annual Review of Economics* 6, 103–128.
- 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.
- SHANNON, C. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423, 623–656.
- SIMS, C. (2003). Implications of rational inattention. *Journal of Monetary Economics* 50, 665–690.
- (2006). Rational inattention: beyond the linear-quadratic case. *American Economic Review* 96, 158–163.
- THOMSON, W. (1979). Eliciting production possibilities from a well-informed manager. *Journal of Economic Theory* 20, 360–380.
- ZHONG, W. (2017). Optimal dynamic information acquisition. *Working Paper*.