NONDISTORTIONARY BELIEF ELICITATION

MARCIN PĘSKI AND COLIN STEWART

UNIVERSITY OF TORONTO

ABSTRACT. A researcher wants to ask a decision-maker about a belief related to a choice the decision-maker made; examples include eliciting confidence or cognitive uncertainty. When can the researcher provide incentives for the decision-maker to report her belief truthfully without distorting her choice? We identify necessary and sufficient conditions for nondistortionary elicitation and fully characterize all incentivizable questions in three canonical classes of problems. For these problems, we show how to elicit beliefs using variants of the Becker-DeGroot-Marschak mechanism.

Keywords: belief elicitation, cognitive uncertainty, experimental design

1. INTRODUCTION

Experimentalists frequently elicit subjects' beliefs about their choices. Examples include the following: (i) The subject chooses an action with a payoff that depends on an unknown state of the world. The researcher asks what probability the subject assigns to his action being correct, i.e, maximizing the *ex post* payoff. (See, e.g., Coffman (2014).) (ii) The subject provides a guess of some quantity and receives a reward according to how close his guess is to the true quantity. The researcher asks him the likelihood that his guess is within some fixed amount x of the correct value. (See, e.g., Enke and Graeber (2023).) (iii) The subject takes a test consisting of a number of multiple choice questions, with a reward for each correct answer. The researcher asks him about the probability that he received a particular score. (See, e.g., Gillen, Snowberg, and Yariv (2019).)

To ensure subjects' reported beliefs are reliable, researchers typically provide incentives that make truthful reporting uniquely optimal.[1] However, when the belief to be elicited is tied to an action choice, doing so could distort the incentives governing that

---

[1]See Healy and Leo (2024) for a discussion of incentivized vs. unincentivized belief elicitation.

choice. To take a simple example, suppose the subject must answer a multiple-choice question and then is asked the probability that they gave the correct answer. Suppose moreover that the subject is rewarded at the belief elicitation stage with a payment that is increasing in the probability the subject assigns to the true event (namely, whether their answer was correct or not). Then a subject who is not confident about the correct answer but is confident that one of the answers is incorrect may be able to increase their overall expected payment by choosing the obviously incorrect answer and then reporting a high probability that it is not correct, thereby obtaining a high expected payoff at the belief elicitation stage.[2]

We ask for what questions truthful reports can be incentivized without distorting the incentives in the original decision problem. For those questions that can be incentivized in this way, we construct simple mechanisms to do so.

Designing payments for belief elicitation that do not distort the incentives in the original problem allows the researcher to honestly tell the subject that she will maximize her expected payment by choosing the action she believes is optimal in the decision problem and then reporting her belief truthfully. [3] Such instructions should minimize any attempts by the subject to distort her behavior, even in cases where profitable distortions are not obvious.

We introduce a model combining a general decision problem with a belief elicitation stage. The model allows us to consider a wide variety of belief elicitation questions that can depend on the action a chosen in the decision problem. Formally, a question asks the subject for his subjective expectation of a function $X(a; \theta)$ (according to his belief about the unknown state θ). For example, $X(a, \theta)$ could be an indicator function for the event that a is an optimal action in state θ, which corresponds to asking how likely a is to be optimal *ex post*. Alternatively, $X(a; \theta)$ could be equal to the utility function $u(a; \theta)$ in the decision problem, which corresponds to asking what the subjective expected utility is in the decision problem.

---

[2]Similar problems arise outside of the experimental context. For example, when a broker (the "subject") chooses an investment portfolio on behalf a client (the "researcher"), the client might ask about the probability that the returns will exceed a certain threshold. Shareholders might ask a CEO about the expected profit from a particular strategy she is choosing to pursue. A builder might ask a contractor about the likelihood of finishing a project by a given date.

[3]One alternative approach would be not to inform subjects about the belief elicitation stage until after they have chosen an action. This approach is unlikely to be effective in experiments with repeated choices and more than one instance of belief elicitation. In any case, we would view this approach as misleading if the experimental design relies on the subject believing that they can maximize their earnings by treating the decision problem in isolation.

We say that a question is *incentivizable* if there exists a payment scheme at the belief elicitation stage for which (i) truthfully reporting the expectation of $X(a;\theta)$ is always the unique maximizer of the subject's expected payment, and (ii) the incentives in the decision problem are not distorted, meaning that, for any belief the subject may have about θ, the set of optimal actions in the decision problem is the same as in the combined problem that includes the belief elicitation stage.

We first identify questions that are incentivizable regardless of the decision problem; we refer to these questions as being aligned with the utility $u(a;\theta)$ in the decision problem. Alignment allows for all questions of the form $X(a;\theta) = u(a;\theta) + d(\theta)$, as well as all invertible affine transformations of such questions with parameters that may depend on the action a. Examples include asking the subject about the payoff she expects to receive in the decision problem, or asking her willingness to pay to have her action replaced with an *ex post* optimal one.[4] On the other hand, a question that asks the subject about the probability that her choice is *ex post* optimal does not generally take this form (and indeed is, in many problems, not incentivizable). A researcher interested in eliciting a measure of cognitive uncertainty (Enke and Graeber, 2023) may therefore do better to ask about the subject's expected payoff relative to the optimum rather than the probability that their action is optimal.

For questions that are aligned with the utility, we provide a simple construction of payments satisfying both of our incentivizability criteria. This construction is based on the classic Becker-DeGroot-Marschak method. One can first normalize $X(a;\theta)$ to lie in $[0,1]$, and then elicit the value of $y \in [0,1]$ at which the subject is indifferent between winning a prize with probability y and winning it with probability $X(a;\theta)$. The expected reward from this mechanism is monotone in the expectation of $X(a;\theta)$; alignment with $u(a;\theta)$ ensures that it is also monotone with respect to the expected utility in the decision problem. A similar construction applies to other questions that we show are incentivizable in some decision problems.

Which other questions are incentivizable, if any, depends on the structure of the decision problem. A particularly important role is played by what we call the "adjacency graph." Two actions are adjacent if there is some belief at which they are both optimal and no other action is. We show that each adjacency places restrictions on

---

[4]The latter question directly extends a simpler one used by Hu (2023) to elicit whether subjects are uncertain about the optimality of their choices.

how the questions following the adjacent actions are related to one another. Problems with more adjacencies therefore tend to involve stronger restrictions on which questions are incentivizable.

We fully characterize the set of incentivizable questions in three canonical classes of decision problems that differ in the structure of their adjacency graphs: adjacency trees, complete adjacency, and product adjacency. (For complete and product adjacency, we also make some mild richness assumptions regarding linear independence of payoffs.)

Adjacency trees naturally arise in many problems with ordered states and actions that are monotone in beliefs. For example, if states and actions are real numbers and the subject incurs a quadratic loss based on the distance between her action and the state, then the adjacency graph forms a line.[5] Because there are so few adjacencies, this case is, in a sense, the most permissive in terms of which questions are incentivizable. In particular, alignment with the utility on the full set of actions is no longer necessary: it suffices for the question to be "piecewise aligned," meaning, in this case, that for each pair of adjacent actions it is aligned with the utility (but with the parameters governing the alignment possibly differing across pairs).

Complete adjacency graphs naturally arise in problems in which the decision-maker chooses an action to match an unknown state and receives a payoff based on whether or not they succeed, as in a multiple-choice problem where the state corresponds to which answer is correct and the subject receives a payment for a correct answer. Relative to adjacency trees, the presence of cycles in the adjacency graph imposes additional restrictions on which questions can be incentivized. In these problems, only questions that are aligned with the utility (in the sense described above) are incentivizable.

Product adjacencies arise when the decision problem comprises a number of separate tasks with complete adjacency graphs and the subject's expected reward is a sum of rewards across these tasks. For example, the subject may complete a multiple-question test and receive a payment proportional to her score. More generally, the researcher may ask the subject about her choices in an experiment with a sequence of tasks, one of which is randomly chosen to be rewarded. The adjacency graph has a special structure: two actions are adjacent if only if they differ on a single task. In this case, a question is incentivizable if and only if it is aligned with some weighted

---

[5]A typical example is when the decision problem is itself a belief elicitation problem—such as one of belief updating—with the states representing objective probabilities known to the researcher.
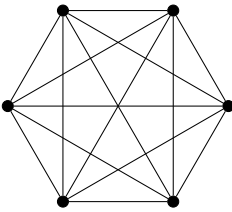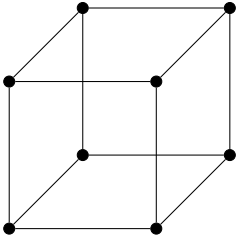
	Tree	Complete graph	Product structure
Adjacency graph			
Examples	Cognitive uncertainty (Enke and Graeber, 2023), monotone one-dimensional problem	Multiple choice question, match-the-state problem	Random problem selection, multiple-question test
Condition	Piecewise aligned	Aligned	Weighted aligned

FIGURE 1. Summary of main results. Nodes of the adjacency graph are actions, with edges indicating adjacencies.

sum of the utilities in the various tasks (but not necessarily aligned with the overall utility in the decision problem). Thus, for example, a question that asks the subject about the likelihood that her score is above some fixed cutoff is not incentivizable, while a question that asks about the expected improvement in her score across two parts of the test is.

Figure 1 illustrates the graphs and summarizes the results for all three classes of decision problems.

Two related features distinguish our approach from previous work on belief elicitation. First, the researcher asks the subject only to report a single number.[6] Second, the quantity of interest to the researcher—as described by the question $X(a;\theta)$—depends nontrivially on the subject's choice of action in the decision problem. In the absence of either of these features, any question $X(a;\theta)$ is incentivizable using standard methods. For example, if the researcher could ask the subject to report his entire belief, it would be enough to incentivize truthful reports and randomly reward the subject either for his choice in the decision problem or for his reported belief. From a practical perspective, however, this approach could be burdensome for subjects if there are more than a few states to report on; if the researcher is only interested in a one-dimensional statistic of the belief, asking about it directly could save time and reduce noise in the reports.

---

[6]We discuss in Section 9.6 how to extend our methods to multiple questions.

1.1. **Related literature: experiments.** Elicitation of subjects' beliefs about their own performance on a task is common in the literature on self-confidence, over-confidence, and motivated reasoning. In a typical experiment, subjects complete a multiple-choice test, such as an IQ test. Following the test, they are asked about their relative or absolute performance (or both). For instance, in Zimmermann (2020) and Benoît, Dubra, and Moore (2015), subjects are asked about the likelihood of being above the median; in Gillen, Snowberg, and Yariv (2019), subjects are asked how many questions they answered correctly and where they lie in the distribution and, in a different task, to report confidence in their decision on a qualitative scale; in Burks et al. (2013), subjects are asked what quintile their performance lies in.[7] In most of these experiments, the belief elicitation is incentivized, either through a quadratic scoring rule (e.g., Zimmermann, 2020; Clark and Friesen, 2009), a BDM mechanism (e.g., Benoît, Dubra, and Moore, 2015), or a payment for the correct answer (e.g., Burks et al., 2013; Serra-Garcia and Gneezy, 2021). These incentives are typically distortionary, and our results suggest that nondistortionary elicitation is, in most cases, not possible for the confidence questions that have been asked, but would be possible for alternative confidence questions based directly on expected payoffs.

Elicitation questions of this form are particularly common in the experimental literature on gender and discrimination (e.g., Hoff and Pandey, 2006; Coffman, 2014; Bordalo et al., 2019; Exley and Nielsen, 2024). In the seminal paper of Niederle and Vesterlund (2007), subjects complete an arithmetic task and guess their own rank in a group of four, receiving a bonus if their guess is correct. In the competition treatment, they are paid for their performance in main task only if they have the highest score in their group. The belief elicitation bonus creates distortionary incentives: a subject who believes he is unlikely to be the top in the group may be able to increase his expected payment by performing badly in the main task and guessing that he has the lowest rank, thereby increasing his likelihood of getting the belief elicitation bonus.

The problem we study is motivated in part by the recent literature on cognitive uncertainty initiated by Enke and Graeber (2023) (see also Amelio, 2022; Arts, Ong, and Qiu, 2024; Xiang, Graeber, Enke, and Gershman, 2021; Clippel, Moscariello, Ortoleva, and Rozen, 2024). In each of these papers, subjects' cognitive uncertainty

---

[7]Other examples include Butler and Loomes (2007), Blavatskyy (2009), Clark and Friesen (2009), Eil and Rao (2011), Ertac (2011), Ortoleva and Snowberg (2015), Serra-Garcia and Gneezy (2021), Möbius, Niederle, Niehaus, and Rosenblat (2022), and Abdellaoui, Bleichrodt, and Gutierrez (2024), among many others. For elicitation of confidence in the psychological literature, see, for example, Liberman (2004), Moore and Healy (2008), and Hoffrage (2016).

is elicited using unincentivized questions. Hu (2023) is the first paper we are aware of that provides strict nondistortionary incentives for subjects to reveal whether they are uncertain about their decision in a complex choice task. His mechanism is essentially a simplified version of the Becker-DeGroot-Marschak mechanism we employ; in his, subjects make a binary choice of whether to pay a cost to have some chance that their action can be replaced with the optimal one.

Several papers explicitly address the possibility of distortions due to belief elicitation incentives in particular experiments. Möbius, Niederle, Niehaus, and Rosenblat (2022) describe how their belief elicitation mechanism is designed with the intention of preserving incentives in their main task. Clark and Friesen (2009) note that the quadratic scoring rule they employ for belief elicitation introduces distortionary incentives, which they try to mitigate by making the incentives in the belief elicitation problem relatively weak. Coffman (2014) elicits, for each question in a multiple-choice test, subjects' beliefs in the likelihood that their answer is correct; the BDM mechanism she uses is nondistortionary for this task.

In game theory experiments, subjects are sometimes asked about their beliefs regarding other players' actions. In an extensive-form game, treating the other players' strategies as an unknown state, such a question may depend on the subject's action. Nyarko and Schotter (2002) elicit beliefs about the action of the other player in simultaneous-move 2x2 games. They note the potential for a different kind of distortion from the one we study: subjects may want to coordinate on predictable action profiles to increase their payments from the belief elicitation mechanism. Aoyagi, Fréchette, and Yuksel (2024) elicit beliefs about the stage-game actions of the other player in a repeated prisoners' dilemma; as in our model, this question depends on the subject's actions insofar as the other player's action is a function of the history of play. Rutström and Wilcox (2009) compare incentivized and unincentivized belief elicitation in a repeated game. They find that belief elicitation incentives affect the play of the game, which may be due more to cognitive factors than to distortion of incentives. Some experiments with sequential-move games elicit beliefs about responses to all possible actions, not only the action actually chosen by the subject, eliminating any possibility of distortions (e.g., Bellemare, Kröger, and Van Soest, 2008; Trautmann and Kuilen, 2015; Fehr, Powell, and Wilkening, 2021).

Belief elicitation is also common in field experiments. In some cases, researchers elicit the subject's willingness-to-pay (WTP) to participate in the experiment (e.g., Ashraf, Berry, and Shapiro, 2010; Dupas, 2014; Dizon-Ross, 2019; Belzil, Maurel,

and Sidibé, 2021). Since the WTP generally depends on the subject's behavior in the main task, this question is action-dependent as in our model. Although some of the observability assumptions of our model may be violated (including observability of actions, states, and payoffs), our results nonetheless imply that eliciting WTP is incentivizable with a BDM mechanism.

The binarized scoring rule introduced by Smith (1961)—and popularized by Hossain and Okui (2013)—has become a common choice for eliciting beliefs in experiments. Danz, Vesterlund, and Wilson (2022) find that subjects report more accurate beliefs when they are told that reporting truthfully will maximize the payment they can expect to receive than when the payments in the binarized scoring rule are described explicitly. In keeping with this finding, we would expect to see less distortion in behavior in belief elicitation settings like ours if subjects are instructed that the incentives are designed to ensure they can maximize their expected payment by treating each task in isolation and choosing what they believe to be optimal; such instructions can be honestly provided only if the belief elicitation question is incentivizable.

1.2. **Related literature: theory.** Belief elicitation has been widely studied and used in both theory and experiments (see Schlag, Tremewan, and Van der Weele (2015), Charness, Gneezy, and Rasocha (2021), Haaland, Roth, and Wohlfart (2023), and Healy and Leo (2024) for surveys). We are not the first to observe that incentivized belief elicitation can distort other decisions. Chambers and Lambert (2021) and Healy and Leo (2024) discuss the possibility that a subject would purposefully fail a test to increase the payment from belief elicitation about the likelihood of passing. Blanco, Engelmann, Koch, and Normann (2010) (see also Schotter and Trevino (2014)) find evidence that, in some problems, subjects who are paid for both a choice in a game and a reported belief take advantage of hedging opportunities, distorting either choice (or both). We implicitly assume that subjects are randomly paid either for the main task or the reported belief, eliminating such risk-hedging opportunities (while retaining other interactions between the action choice and the belief report).

Chassang, Padró I Miquel, and Snowberg (2012) develop a theory of elicitation of beliefs about returns to an investment that depends on unobservable effort. In their model, correct elicitation of beliefs requires that the effort choice is not distorted. They show that a willingness-to-pay method can be used to identify returns without distorting incentives in the effort choice.

The closest theoretical work to this proposal is that of Lambert, Pennock, and Shoham (2008) and Lambert (2019), which ask which properties of distributions can

be elicited. Our model shares the feature that the belief elicitation question does not ask the decision-maker to report her entire belief. We sidestep their question of elicitability by restricting attention to questions that always correspond to elicitable properties, and we add the condition that the elicitation must not distort the decisions in the main decision problem.

Azrieli, Chambers, and Healy (2018) study incentives in a sequence of tasks and find that paying for a randomly selected problem is the only incentive-compatible mechanism when allowing for a general class of preferences. In their model, the sequence of tasks is exogenously given, whereas in ours the belief elicitation task depends on the subjects' choice in the main task. Random selection for payment is therefore not sufficient to ensure incentive compatibility.

2. MODEL

A subject (he) chooses an action and then faces a belief elicitation problem posed by a researcher (she) that may depend on the action he chose.

A *decision problem* consists of a tuple (Θ, A, u), where Θ is a finite set of states of the world, A is a finite set of actions, and $u : A \longrightarrow \mathbb{R}^{\Theta}$ is a utility function specifying, for each action, the vector of payoffs across all states. We write $u(a; \theta)$ for the θ-coordinate of the vector $\mathsf{u}(a)$. (Note that we use a distinct font for vectors to distinguish, for example, between the vector $\mathsf{u}(a)$ and the function u.) For each belief $p \in \Delta(\Theta)$, let $A(p) = \arg\max_{a \in A} \sum_{\theta} p(\theta) u(a; \theta)$ denote the set of optimal actions at p; we refer to actions in $A(p)$ as u-optimal. For simplicity, we assume that (i) there are no redundant actions, i.e., no a and a' such that $\mathsf{u}(a) = \mathsf{u}(a')$, and (ii) there are no weakly dominated actions, i.e., for each a, there exists some p such that $A(p) = \{a\}$.

After choosing an action a, the subject faces a *question* $\mathsf{X}(a) \in \mathbb{R}^{\Theta}$ about his belief chosen according to a *question profile* $X : A \longrightarrow \mathbb{R}^{\Theta}$, with θ-coordinate $X(a; \theta)$. We interpret the question $\mathsf{X}(a)$ as asking the DM to report his subjective expected value $\mathbb{E}_p X(a; \cdot)$ given the action a that he chose in the first stage and his belief p. The dependence of X on a allows for the possibility that the researcher seeks information about the subject's belief that is related to the chosen action.

The following examples illustrate how this formulation allows for considerable flexibility in what the subject is asked to report:

(1) The question "how likely is it that the chosen action a is optimal *ex post*?" corresponds to

$$X(a;\theta) = \mathbb{1}\{a \in \arg\max_{b \in A} u(b;\theta)\}.$$

Then $\mathbb{E}_p X(a;\cdot)$ is the subjective probability of the chosen action being u-optimal *ex post*.

(2) The question "what is the expected regret from the chosen action?" corresponds to

$$X(a;\theta) = \max_{b \in A} u(b;\theta) - u(a;\theta). \tag{1}$$

(3) The question "how likely is it that your guess a is within x of the true value θ?" corresponds to

$$X(a;\theta) = \mathbb{1}\{|a - \theta| \leq x\}. \tag{2}$$

The subject announces a report $r \in \mathbb{R}$ and is given a reward that may depend on his report, his action a, and the realized state θ. His overall payoff—including the payoff from the decision problem in the first stage—is given by a bounded function $V : \mathbb{R} \times A \times \Theta \longrightarrow [0,1]$ (which we normalize to the unit interval for convenience), the first argument of which is the subject's report r. We refer to V as an *elicitation method*.

For simplicity, we do not include the payoff $u(a;\theta)$ explicitly in the elicitation method; adding it would not change anything as the payoff from the elicitation problem can be adjusted accordingly so as to give the same overall payoff. In the applications we have in mind, the researcher pays the subject for the decision problem with some fixed probability $\alpha \in (0,1)$ and for the belief elicitation problem with the remaining probability $1 - \alpha$. The elicitation method V therefore takes the form $\alpha u(a;\theta) + (1 - \alpha)V_0(r,a,\theta)$ for some V_0, where V_0 is the payoff from the belief elicitation mechanism. As is standard in the recent literature on belief elicitation, we implicitly view V_0 as the probability of winning a fixed prize to avoid any influence of risk preferences on the reported belief.

Definition 1. A question profile X is *incentivizable* if there exists an elicitation method V such that, for every $p \in \Delta(\Theta)$,

$$\arg\max_{(r,a)} \mathbb{E}_p V(r,a,\cdot) = \{(\mathbb{E}_p X(a;\cdot),a) : a \in A(p)\}. \tag{3}$$

Any V satisfying this condition *incentivizes* X.

If, given p, (r, a) maximizes $\mathbb{E}_p V(r, a, \cdot)$, we say that a is V-optimal.[8]

Incentivizability combines two requirements of the elicitation method. First, the payoffs at the belief elicitation stage must not distort the subject's action choices in the decision problem in the sense that the set of actions a he optimally chooses in the overall problem with payoffs V is the same as in the original decision problem with payoffs u; in other words, u-optimality and V-optimality are equivalent.[9] Second, the subject must have strict incentives to report her true subjective expectation of $\mathbb{E}_p X(a; \cdot)$ given her action choice a. In an experiment, for any question profile that cannot be incentivized, no matter how the researcher designs the incentives at the belief elicitation stage, she cannot honestly tell subjects that they will maximize their earnings by considering the decision problem in isolation and by reporting beliefs truthfully.

Our formulation imposes two substantive restrictions on the belief elicitation problem. First, the subject is asked to report only a one-dimensional statistic of the belief rather than, say, a full probability distribution. In practice, collecting more complicated information about beliefs quickly becomes impractical beyond a small number of states. If, however, the full probability distribution could be elicited, our problem would reduce to a standard belief elicitation problem since there would be no need to make the question or incentives dependent on the action chosen in the decision problem. Second, the elicited belief is based on the expectation of some question $X(a; \cdot)$. While this formulation captures many relevant cases, in principle, the researcher may want to elicit other properties of the distribution for which our approach may not apply. We discuss in Section 9 what changes if either of these restrictions is relaxed.

The model makes two important implicit assumptions. First, the action, state, and utility function are all observed by the researcher. This assumption typically holds in lab experiments. Second, the action does not directly affect the state or the subject's belief about the state. We discuss both of these assumptions in Section 9.

3. Overview for practitioners

As an illustration of our results, consider an experiment in which the subject makes a choice under uncertainty. The experimenter then wishes to ask the subject about

---

[8]We refer to reports r or pairs (a, r) that maximize $\mathbb{E}_p V(r, a, \cdot)$ simply as "optimal" since there is no risk of confusion with u-optimality.

[9]A researcher might be content with the weaker condition that the left-hand side of (3) is a subset of the right-hand side, ensuring that any V-optimal action is also u-optimal. Since each action is u-optimal on a closed set of beliefs and multiplicity occurs only on the boundaries of these sets, and similarly for V-optimality, this requirement is equivalent to incentivizability as we define it.

his confidence in his decision. To do so, she could ask the subject how likely he thinks it is that his decision is optimal *ex post*, or in other words, how likely it is that if he learns the true state of the world he will not want to change his action; call this the "ex post optimality question." Alternatively, she could ask how much he would be willing to pay to have his action replaced with the *ex post* optimal one; call this the "expected regret question." The difference between these questions is that, in the latter, likelihoods of states in which the action is suboptimal must be weighted according to the loss in payoff that occurs in those states. (Other questions could be used to measure confidence, but we will focus on these two for simplicity.)

Whichever confidence question the experimenter uses, to elicit reliable answers, she would like to provide incentives that make it uniquely optimal for the subject to report the answer truthfully. In addition, she does not want these incentives to contaminate the original decision problem. For example, it should not be that the subject can increase his total expected payment by making a choice that he knows is suboptimal ex post in the main problem, leading to a larger reward at the belief elicitation stage when he correctly reports low confidence in his choice. The question if incentivizable if there exists a payment scheme that satisfies both of these criteria.

It turns out that the ex post optimality question is only incentivizable if the original decision problem has a specific structure. For example, suppose the decision problem consists of a multiple-choice question, with the subject receiving a fixed reward for a correct answer. Formally, the set of actions A, which describe possible answers, is equal to the set of states Θ, which describe the correct answer; the subject's answer is correct if $a = \theta$. The payoff in the decision problem is

$$u(a, \theta) = \begin{cases} R & \text{if } a = \theta, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $R > 0$. In this case, the ex post optimality question amounts to asking the subject how likely he thinks it is that the answer he gave is correct. Notice that the probability that the answer a is correct is equal to the expectation of $u(a, \theta)/R$. Thus the ex post optimality question is described by $X(a; \theta) = u(a, \theta)/R$.

Proposition 1 below implies that this question is incentivizable in this decision problem. Moreover, it can be incentivized using a standard BDM mechanism to elicit the probability equivalent of the event $\theta = a$ for the action a chosen by the subject. In this mechanism, the subject is asked to report the probability r that he assigns to the state $\theta = a$. A random number $z \sim U[0, 1]$ is drawn. If $r \geq z$, the subject wins

a fixed reward $R' > 0$ if indeed $\theta = a$; otherwise, the subject wins the reward with probability z.

As is well known, the BDM mechanism makes it optimal for the subject to report his true belief q. To check whether it distorts incentives in the original decision problem, consider the payoff the subject receives from the mechanism. Given that he reports $r = q$, his subjective expected payoff from the mechanism is

$$R' \left(\int_0^q q dz + \int_q^1 z dz \right) = \frac{R'}{2} \left(1 + q^2 \right). \tag{5}$$

Notice that this expression is increasing in q. Thus the subject maximizes his expected payment in the mechanism by choosing the action a that maximizes his subjective probability that $\theta = a$—which is exactly the action that maximizes his expected payoff in the original decision problem. Therefore, the optimal action in the problem that combines the original decision problem and the belief elicitation mechanism is the same as when facing the original problem in isolation, as needed for incentivizability.

This argument relies on two key features of the mechanism. First, the subject's expected payoff from the mechanism is increasing in q. More generally, this value should be monotone in the subjective expectation of X. The use of the BDM mechanism ensures that this property holds for each a.[10] Moreover, the expected payoff from the mechanism depends only on q, not on the choice of action a. More generally, the use of BDM ensures that this property holds after some suitable renormalization of each $X(a; \cdot)$ by a non-zero affine transformation.

Second, the original decision problem is such that the optimal action is the one that maximizes q; this feature results from X being proportional to u. More generally, following a suitable renormalization, the optimal action a in the original decision problem should be the one that maximizes the expectation of $X(a; \cdot)$. This property holds under a condition that we call "alignment" that extends beyond proportionality of X and u to a larger class of transformations.

For the ex post optimality question, the latter condition fails in many decision problems. Examples 1 and 3 describe two decision problems in which the utility function is more complicated than the one above and is not aligned with X. In these examples, our necessary conditions imply that the ex post optimality question cannot be incentivized.

---

[10]This property does not necessarily hold for other mechanisms that incentivize truthful reports. For example, quadratic binarized scoring rule is not monotone.

On the other hand, the expected regret question is *always* aligned with the utility: $-X$ differs from u by a constant that depends only on θ. This question is therefore incentivizable in every decision problem by using a BDM mechanism. In this case, the first step is to take an affine transformation of X to transform it to a question $\hat{X}$ that takes values in $[0, 1]$ and is increasing as u increases. The random number $z \sim U[0, 1]$ is then compared to the subject's report r; if $z > r$, the subject wins the reward R' with probability z, and if $r \geq z$, the subject wins the reward R' with probability $\hat{X}(a, \theta)$.

Now suppose the decision problem is a test comprising I multiple choice questions, with the subject receiving a reward for each question he answers correctly. Formally, actions take the form $a = (a_i, \ldots, a_I) \in A^I$ and states the form $\theta = (\theta_1, \ldots, \theta_I) \in \Theta^I$, with $A = \Theta$. The payoff in the decision problem is $U(a, \theta) = \frac{1}{I} \sum_{i=1}^{I} u(a_i, \theta_i)$, where u is defined by (4).

Following the test, the experimenter would like to learn some statistic of the subject's belief regarding his performance. For example, the experimenter may want to elicit the subject's subjective expected score on the test. This corresponds to the question $X(a; \theta) = U(a, \theta)/R$, which is incentivizable using a BDM mechanism in essentially the same way as in the case of the expected regret question above.

On the other hand, our results imply that the experimenter cannot incentivize the subject to truthfully report the probability he assigns to achieving a score above a fixed threshold $z \in \{0, \ldots, I - 1\}$ without distorting incentives in the decision problem. For example, suppose the experimenter uses a BDM mechanism to elicit this belief. Because the value associated with this mechanism is increasing in the belief, the subject maximizes his payoff at the belief elicitation stage by maximizing the probability that his score exceeds z. This objective can be in conflict with maximizing the probability of answering each question correctly if the subject does not believe θ_i is independent across i. For a particularly simple (if unrealistic) example, suppose the test asks the same question four times and there are four possible answers (that is, $I = |A| = 4$), and take $z = 0$. The subject can guarantee a score of 1 by providing a different answer to each question; since $1 > z$, this strategy combined with a reported belief that his score exceeds z with probability 1 gives the largest possible payoff at the belief elicitation stage. The strategy that maximizes the subject's expected payoff in the decision problem alone is to provide the same answer—the one that is most likely correct—to all four multiple choice questions. If the subject is not sufficiently

confident about which answer is correct, the former strategy yields a higher overall expected payoff than the latter does.

Some more complicated belief elicitation questions are incentivizable. Suppose the experimenter wants to elicit the subject's belief about his expected improvement across two parts of the test. For example, take I to be even and let

$$X(a; \theta) = \sum_{i=I/2+1}^{I} \mathbb{1}(a_i = \theta_i) - \sum_{i=1}^{I/2} \mathbb{1}(a_i = \theta_i).$$

This question corresponds to asking the subject his expectation of the increase in his raw score from the first half of the test to the second. The previous logic does not apply directly here because increasing $u(a_i, \theta_i)$ increases X for $i > I/2$ but decreases it for $i \leq I/2$.

And yet this question is incentivizable using a BDM mechanism, with the caveat that the incentives in the belief elicitation task cannot be too strong relative to those in the decision problem. To implement this mechanism, we first rescale X according to $\hat{X}(a; \theta) = 1/2 + X(a; \theta)/I$; doing so ensures that it takes values in $[0, 1]$. By (5), reporting the expectation of X truthfully gives an expected payoff of $R'(1 + \mathbb{E}_p[\hat{X}(a, \cdot)]^2)/2$.

Now suppose the subject is rewarded for the decision problem with probability α and for the belief elicitation problem with the remaining probability $1 - \alpha$, where $\alpha \in (0, 1)$. Notice that it is possible for the subject to increase his expected payment from the belief elicitation mechanism by choosing an action profile that is suboptimal in the decision problem, for example by intentionally reducing his expected score in the first half of the test to increase his expected improvement. However, if he chooses an action profile that lowers his expected payoff in the decision problem by some $\Delta > 0$ relative to the optimal profile a^*, doing so will change the value of $\hat{X}$ by at most Δ, and therefore increases the expected payoff at the belief elicitation stage by at most a fixed multiple of Δ. If the incentives in the decision problem are strong enough relative to those at the belief elicitation stage, it is therefore optimal for the subject to choose a^*. In fact, one can show that as long as $\alpha R \geq (1 - \alpha)R'$, there is no distortion in incentives in the decision problem.

The same logic applies to all questions X that are proportional to linear combinations of the payoffs $u(a_i, \theta_i)$, and to much more general decision problems in which the payoff is the sum of payoffs across i, with the action sets, state spaces, and payoffs possibly varying with i.

4. Sufficient conditions

We begin by identifying simple sufficient conditions under which a question profile X is incentivizable. In Section 7, we show that these conditions are also necessary in some natural applications.

Definition 2. Questions $\mathsf{X}, \mathsf{Y} \in \mathbb{R}^\Theta$ are *equivalent* if there exist $\gamma \in \mathbb{R} \setminus \{0\}$ and $\kappa \in \mathbb{R}$ such that

$$\mathsf{X} = \gamma \mathsf{Y} + \kappa \mathbf{1},$$

where $\mathbf{1} \in \mathbb{R}^\Theta$ is the vector of all ones. We say that question profiles X and Y are equivalent if $\mathsf{X}(a)$ and $\mathsf{Y}(a)$ are equivalent for all $a \in A$.

Given two equivalent questions, for any given belief, knowing the expectation of one of them is sufficient to calculate the expectation of the other. Switching from one question to the other therefore amounts to shifting and rescaling the decision-maker's reports, and does not affect the information the researcher would obtain from truthful responses.

Definition 3. A question profile X is *aligned with u on $B \subseteq A$* if there exists $\mathsf{d} \in \mathbb{R}^\Theta$ such that either $\mathsf{X}(a)$ is equivalent to the question $\mathsf{u}(a) + \mathsf{d}$ for all $a \in B$, or $\mathsf{X}(a)$ is equivalent to the question d for all $a \in B$. We say that X is *nontrivially aligned with u on B* in the former case, and *trivially* so in the latter case. If $B = A$, we say simply that X is *aligned with u* (and similarly with the (non)-trivial qualifier).

Relative to the question profile $X = u$ that asks the subject about the expected utility from his chosen action, questions aligned with u allow for three changes. First, a vector d may be added to payoffs. Since d is independent of a, this change has no effect on the u-optimal action. Then, for each a, the question $\mathsf{X}(a)$ can be rescaled by a (nonzero) constant $\gamma(a)$ and translated by another constant $\kappa(a)$ uniformly across θ. These changes make each question $\mathsf{X}(a)$ equivalent to the question $\mathsf{u}(a) + \mathsf{d}$. The case of trivial alignment can be viewed as a limit of these operations as d is scaled up and $\gamma(a)$ scaled down by the same constant, causing the $\mathsf{u}(a)$ term to vanish.

Note that for X aligned with u, the parameters γ, κ, and d are not uniquely determined in general.

Proposition 1. *If X is aligned with u, then it is incentivizable.*

Proofs omitted from the main text may be found in the appendix.

The proposition indicates that alignment with u is sufficient for incentivizability. The proof proceeds by construction using a standard Becker-DeGroot-Marschak

(BDM) mechanism.[11] The idea is to first replace each question with an equivalent one to make it of the form $\mathsf{X}(a) = \mathsf{u}(a) + \mathsf{d}$ for all a, or $\mathsf{X}(a) = \mathsf{d}$ for all a. Let $[L, M]$ be an interval containing every value of $X(a; \theta)$. After learning the subject's report, r, of his expectation of $\mathsf{X}(a)$, the researcher draws a number x uniformly from $[L, M]$. The subject receives a fixed prize with probability $(X(a; \theta) - L)/(M - L)$ if $r > x$, and with probability $(x - L)/(M - L)$ otherwise. By standard arguments, this mechanism provides strict incentives for the subject to truthfully report his expectation of $\mathsf{X}(a)$. Moreover, given that the subject reports truthfully, his expected value from the mechanism is an increasing function of his expectation of $\mathsf{X}(a)$. If $\mathsf{X}(a) = \mathsf{u}(a) + \mathsf{d}$, it follows that, to maximize the expected payoff from the mechanism, the subject should choose an action a that maximizes the expectation of $\mathsf{u}(a)$, as needed for incentivizability. If $\mathsf{X}(a) = \mathsf{d}$, the expected payoff from the mechanism does not depend on the action choice; an overall payoff V equal to u plus the payoff obtained from the BDM mechanism ensures incentivizability.

For a researcher interested in eliciting a measure of the subject's confidence in her choice of action, an immediate implication of the proposition, captured in the following corollary, is that it is possible to elicit the subject's expected regret without distorting her decisions.

Corollary 1. *For any decision problem, the question about regret in equation* (1) *is incentivizable, as is any question profile that does not depend on the chosen action.*

5. NECESSARY CONDITIONS

We now identify simple necessary conditions for question profiles to be incentivizable. Since such questions must not distort incentives for the action choice in the decision problem, it is natural to focus on beliefs where the subject is indifferent between two actions. Any distortion that arises at a point of indifference implies that there are also distortions in an open set around that point.

Say that two actions $a, b \in A$ are *adjacent* if there is a belief $p \in \Delta(\Theta)$ such that $A(p) = \{a, b\}$, that is, at belief p, a and b are both u-optimal and there is no other u-optimal action.

---

[11]While all of our constructions employ variants of BDM, these are presumably not the only nondistortionary elicitation methods that incentivize truthful reporting of beliefs for incentivizable question profiles. Identifying the set of such elicitation methods is beyond the scope of this paper.

Our necessary conditions make use of an observation about the structure of the *value of information*

$$V^*(p) = \max_{a,r} \mathbb{E}_p V(a, r, \cdot);$$

$V^*(p)$ is the payoff associated with any optimal pair (a, r) at belief p. Note that $V^*(p)$ is the upper envelope of the affine functions $\mathbb{E}_p V(a, r, \cdot)$ across (a, r), and is therefore convex. Moreover, given beliefs $p \neq p'$, if there exists some (a, r) that is optimal at both p and p', then V^* is affine along the line segment $\overline{pp'}$ connecting p to p'. On the other hand, if for some $\alpha \in (0, 1)$ there is a pair (a, r) that is optimal at $q = \alpha p + (1 - \alpha)p'$ but is not optimal at one of p or p', then the value of information is not affine along $\overline{pp'}$.[12]

Consider an elicitation method V that incentivizes a question profile X. Let p be a belief with u-optimal actions $A(p) = \{a, b\}$. Notice that the values associated with actions a and b must match at belief p: $\mathbb{E}_p V(a, r, \cdot) = \mathbb{E}_p V(b, s, \cdot)$ for the reports $r = \mathbb{E}_p X(a; \cdot)$ and $s = \mathbb{E}_p X(b; \cdot)$. If (a, r) is also optimal at another belief p' with $A(p') = \{a, b\}$, then by the above observation, the value of information is affine along the line segment $\overline{pp'}$, and (a, r) is optimal everywhere along $\overline{pp'}$. Because the values must match everywhere along $\overline{pp'}$, $\mathbb{E}_q V(b, s(q), \cdot)$ is affine along $\overline{pp'}$ for the optimal report $s(q)$ at each belief q. By the above observation, it follows that (b, s) is optimal everywhere on $\overline{pp'}$.

These observations impose strong conditions on how the questions $\mathsf{X}(a)$ and $\mathsf{X}(b)$ relate to one another for adjacent actions a and b. Formalizing this idea requires two additional pieces of notation. First, for any vector $\mathsf{v} \in \mathbb{R}^\Theta$, let $\bar{\mathsf{v}} = \mathsf{v} - \frac{1}{|\Theta|} \sum_{\theta' \in \Theta} v(\theta')\mathbf{1}$ be the projection of v onto the hyperplane of vectors whose coordinates sum to 0. Second, let $\Delta_a^b = \bar{\mathsf{u}}(b) - \bar{\mathsf{u}}(a)$ be the payoff difference vector.

Lemma 1 (Adjacency Lemma). *Suppose X is incentivizable. Then it is aligned with u on $\{a, b\}$ for every pair of adjacent actions a and b. Equivalently, for each pair of adjacent actions a and b, there exist $\rho \in \mathbb{R}$ and $\sigma \in \mathbb{R} \setminus \{0\}$ such that*

$$\bar{\mathsf{X}}(b) = \rho \Delta_a^b + \sigma \bar{\mathsf{X}}(a);$$

if $\bar{\mathsf{X}}(a)$ or $\bar{\mathsf{X}}(b)$ is collinear with Δ_a^b, then we can take $\rho \neq 0$.[13]

---

[12]To see why, first suppose (a, r) is optimal at p and p'. Since V^* is convex, for any convex combination $q = \alpha p + (1 - \alpha)p'$, we have $\mathbb{E}_q V(a, r, \cdot) \leq V^*(q) \leq \alpha V^*(p) + (1 - \alpha)V^*(p') = \alpha \mathbb{E}_p V(a, r, \cdot) + (1 - \alpha)\mathbb{E}_{p'} V(a, r, \cdot) = \mathbb{E}_q V(a, r, \cdot)$, which implies that all inequalities hold with equality, as needed. For the converse, suppose (a, r) is optimal at q but not at p. Then $\alpha V^*(p) + (1 - \alpha)V^*(p') > \alpha \mathbb{E}_p V(a, r, \cdot) + (1 - \alpha)\mathbb{E}_{p'} V(a, r, \cdot) = \mathbb{E}_q V(a, r, \cdot) = V^*(q)$.

[13]Recall that two vectors $\mathsf{u}, \mathsf{v} \in \mathbb{R}^\Theta$ are *collinear* if there exists $\alpha \neq 0$ such that $\mathsf{v} = \alpha \mathsf{u}$.

The Adjacency Lemma provides a key tool in testing whether a question profile is incentivizable: it identifies a restriction on the values of the question at each pair of adjacent actions a and b.

The proof of the lemma has two steps. The first step uses the above observations about the value of information to show that, for adjacent actions a and b and beliefs p and p' at which a and b are u-optimal, if the same report r is optimal at beliefs p and p' following action a, then there must also be a report s that is optimal at both p and p' following action b. For the second step, we employ a linear algebra argument to show that the vector $\mathsf{X}(b)$ must belong to the linear space spanned by the vectors $\mathsf{u}(b) - \mathsf{u}(a)$, $\mathsf{X}(a)$, and $\mathbf{1}$ (the latter because we apply the linear condition to the space of beliefs, which satisfy the condition $\sum p(\theta) = 1$). The result then follows from straightforward algebra.

Adjacency of actions a and b implies that there is a $(|\Theta| - 2)$-dimensional set of beliefs $p \in \Delta(\Theta)$ for which actions a and b are uniquely u-optimal. Since the Adjacency Lemma is based on arguments restricted to this space, it has no bite when $|\Theta| = 2$—indeed, when $|\Theta| = 2$, every question profile is aligned with u on each pair of adjacent actions—and limited bite when $|\Theta| = 3$. We discuss this issue in Section 9.1.

Example 1 (Second-order beliefs)**.** The following example is a slightly simplified version of a belief-updating experiment from Enke and Graeber (2023); the same conclusions apply to their original experiment.

The decision problem involves forecasting a binary event. The action set and state space $A = \Theta = \{0, 1/n, \ldots, 1\}$ consist of (discretized) probabilities that the event occurs, where $n \geq 4$. One can think of θ as the "true" probability given the available information, which is known to the researcher but about which the subject may be uncertain (for example because he has doubts about how to update his beliefs in light of the information he observes). This uncertainty is captured by the belief $p \in \Delta(\Theta)$.

The subject is rewarded more for forecasts that are closer to the state according to the payoff function $u(a; \theta) = -(a - \theta)^2$. The u-optimal actions are those that are closest to $\mathbb{E}_p[\theta]$. The adjacency graph therefore forms a line: a_i and a_j are adjacent if and only if $|a_i - a_j| = 1/n$.

For some $x \in [0, 1/2]$, the researcher wishes to elicit the subject's confidence in his choice by asking how likely he believes it is that his action is within x of the true value of θ, as in equation (2).

To check whether X is incentivizable, we use Lemma 1. Note that the coordinates of the vector $\bar{\mathsf{X}}(a)$ take at most two values. Hence if a and $b = a + 1/n$ are two adjacent actions and $\sigma \neq 0$, then the coordinates of $\bar{\mathsf{X}}(a) - \sigma \bar{\mathsf{X}}(b) \neq 0$ take at most four distinct values. At the same time, for each θ,

$$\Delta_a^b(\theta) = \bar{u}(b; \theta) - \bar{u}(a; \theta) = \frac{2\theta - 1}{n},$$

and hence the coordinates of Δ_a^b take on $n + 1 \geq 5$ different values. By the Adjacency Lemma, since $\bar{\mathsf{X}}(a) - \sigma \bar{\mathsf{X}}(b) \neq 0$ implies $\rho \neq 0$, X is not incentivizable.

This result is not specific to the quadratic payoffs in the decision problem: the same conclusion applies if the payoff is replaced with any strictly proper scoring rule (one for which reporting an action close to the expectation of θ is u-optimal).

Faced with this result, what should the researcher do? One option is to use a different measure of decision confidence that *is* incentivizable. For instance, according to Corollary 1, the expected regret question of equation (1) is incentivizable in every decision problem.

There is a gap between the necessary condition for incentivizability from the Adjacency Lemma —namely, alignment on all adjacent pairs—and the sufficient condition from Proposition 1—namely, alignment on the full set of actions. In the rest of the paper, we show how to close this gap in three canonical classes of decision problems. In one of those classes, alignment on adjacent pairs is sufficient (and necessary); in another, alignment on the full set of actions is necessary (and sufficient); and in the third, a weaker form of alignment on all actions is both necessary and sufficient.

Our approach relies on the following observation. If the decision problem has three actions a, b, and c such that $\{a, b\}$ and $\{b, c\}$ are both adjacent pairs, the restrictions implied by the Adjacency Lemma for these two pairs may interact with each other, leading to additional information about which questions are incentivizable. We therefore look to analyze the restrictions across all adjacent pairs simultaneously.

The *adjacency graph* is the undirected graph with vertices A and edges consisting of the adjacent pairs $\{a, b\}$. Note that, since there are no redundant or dominated actions, the adjacency graph is connected for every decision problem. A basic intuition across the next three sections is that the more edges there are in the adjacency graph, the more powerful are the restrictions imposed by the Adjacency Lemma.

6. ADJACENCY TREES

We first consider the case in which the adjacency graph is a tree (i.e., has no cycles). Trees naturally arise in problems with ordered actions, as in Example 1. Example 2 below describes a simple decision problem in which the adjacency graph forms a star. In this case, as the following result shows, alignment with u on the full set of actions is not necessary for incentivizability.

Theorem 1. *Suppose the adjacency graph is a tree. Then X is incentivizable if and only if it is aligned with u on every pair of adjacent actions.*

That alignment with u on adjacent pairs is necessary for incentivizability holds in general, as indicated by Lemma 1. That it is sufficient follows from Proposition 2 below, which is a generalization of Proposition 1.

Say that action a is *splitting* if removing it from the adjacency graph makes the graph disconnected. A *splitting collection* $\{A_0, \ldots, A_k\}$ is a collection of subsets $A_i \subseteq A$ such that $\bigcup A_i = A$ and, for each i and j, either $A_i \cap A_j = \emptyset$ or $A_i \cap A_j = \{a\}$ for some splitting action a. If the adjacency graph is a tree, a splitting collection is formed by the set $\mathcal{A} = \{A_0, \ldots, A_k\}$ consisting of all pairs of adjacent actions; that is, $\tilde{A} \in \mathcal{A}$ if and only if $\tilde{A} = \{a, b\}$ for some adjacent actions a and b.

Example 2. Suppose the decision problem involves guessing the correct state, with the option of taking a safe action. The action set is $A = \Theta \cup \{a_s\}$, with payoffs $u(\theta, \theta) = 1$, $u(a_s, \theta) = s$ for some $s \in (0, 1)$, and $u(a, \theta) = 0$ for $a \neq \theta, a_s$. If $s \geq 1/2$, then the adjacency graph is a star with action a_s in the centre. Letting $\{B_0, B_1\}$ be any partition of Θ, $\{B_0 \cup \{a_s\}, B_1 \cup \{a_s\}\}$ forms a splitting collection. Another splitting collection is given by the set of pairs $\{\theta, a_s\}$ for all $\theta \in \Theta$.

We say that X is *piecewise aligned with u* if it is aligned with u on each element of some splitting collection. Theorem 1 follows from the Adjacency Lemma together with the following result.

Proposition 2. *If X is piecewise aligned with u, then it is incentivizable.*

To understand the main idea of the proof of this result, consider a binary splitting collection with sets of actions A_0 and A_1 and splitting action a_0. If actions are restricted to either A_i, then by the piecewise alignment assumption, the BDM construction of Proposition 1 can be used to incentivize X. This construction gives rise to a well-defined elicitation method on the full action set if the two methods agree on

$A_0 \cap A_1 = \{a_0\}$. In the proof, we construct a positive affine transformation of one of the elicitation methods that ensures agreement on a_0.

We show that the new elicitation method incentivizes X on the union of the two sets. Indeed, by construction, given any belief p at which an action $a \in A_i$ is u-optimal, no other action in A_i leads to a higher expected payoff from the elicitation method. It remains to show that, similarly, no action $a' \in A_j$ (for $j \neq i$) does better than a at p. To do so, we prove that if p_0 and p_1 are beliefs such that, for each i, some action in A_i is u-optimal at belief p_i, then there exists a convex combination of p_0 and p_1 at which a_0 is u-optimal. Taking one of these beliefs to be p and the other to be a belief at which a' is u-optimal, it follows that there is a belief at which a_0 is u-optimal but not V-optimal, contrary to the way the method was constructed.

7. ADJACENCY CYCLES

When the adjacency graph forms a tree, the incentivizable questions are those that are piecewise aligned with u; the relationship between $\bar{X}(a)$ and $\bar{X}(b)$ described in the Adjacency Lemma is both necessary and sufficient for incentivizability. Cycles in the adjacency graph impose additional restrictions: not only must the relationship in the Adjacency Lemma hold for adjacent actions, it must also be consistent all the way around each cycle.

The additional restrictions cycles impose are related to linear independence relationships among the corresponding utility vectors. To understand how, consider a cycle $(a_0, \ldots, a_n)$ in the adjacency graph (where $a_0 = a_n$). Recall that alignment with u on the adjacent pair $\{a_i, a_{i+1}\}$ implies

$$\bar{X}(a_i) = \rho \Delta^{a_i}_{a_{i+1}} + \sigma \bar{X}(a_{i+1})$$

with $\rho \neq 0$. Iterating these equations around the cycle starting from a_0 leads to

$$\bar{X}(a_0) = \kappa \bar{X}(a_0) + \sum_{i=0}^{n-1} \rho_i \Delta^{a_i}_{a_{i+1}}$$

for some κ and $\rho_i \neq 0$. If the $\Delta^{a_i}_{a_{i+1}}$ terms are linearly independent, it must be that $\kappa \neq 1$, and therefore $\bar{X}(a_0)$ is a linear combination of the vectors $\bar{u}(a_i)$.

A common setting in which adjacency cycles appear is when the adjacency graph is complete. This is the case, for instance, in Example 2 when $s < 1/2$.

The main result of this section shows that, under a mild richness condition on utilities, in decision problems with complete adjacency graphs, a question is incentivizable if and only if it is aligned with u.[14]

Theorem 2. *Suppose the adjacency graph is complete and $|A| \geq 4$. Suppose in addition that for any four distinct actions a, b_0, b_1, b_2, the set of vectors $\{\Delta_a^{b_i}\}_{i=0,1,2}$ is linearly independent. Then X is incentivizable if and only if it is aligned with u.*

In addition to completeness of the adjacency graph, this theorem relies on two assumptions: the set of actions must be sufficiently large and the payoffs from the actions must be sufficiently independent. These assumptions make the Adjacency Lemma particularly powerful. For example, as discussed in Section 5, the Adjacency Lemma only has any bite if there are at least three states of the world, which must be the case if the two assumptions of the theorem hold. These assumptions also exclude decision problems with larger state spaces that can effectively be reduced to problems with three states, such as those in which the payoff from each action is 0 in every state but the first three.

Example 3. Consider a decision problem with at least four states in which the subject is asked to guess the state and receives a reward for guessing correctly that may depend on which state is realized. Thus $A = \Theta$ and

$$u(a; \theta) = r_\theta \cdot \mathbb{1}\{a = \theta\},$$

where $r_\theta > 0$ for all θ. The adjacency graph for this problem is complete: given any two states θ and θ', the actions θ and θ' are the only u-optimal actions at the belief that assigns probability $r_{\theta'}/(r_\theta + r_{\theta'})$ to state θ and all of the remaining probability to state θ'. Note that the linear independence condition of Theorem 2 is satisfied.

Suppose the researcher seeks to elicit the subject's belief about whether she correctly guessed the state, which is described by the question profile

$$X(a; \theta) = \mathbb{1}\{a = \theta\}.$$

Taking $\gamma(a) = 1/r_a$ and $\kappa(a) = 0$ for each a and $\mathsf{d} = 0$, we see that X is aligned with u, and therefore can be incentivized using the BDM construction of Proposition 1.

---

[14]Theorem 2 is related to Proposition 3 of Morris and Ui (2004), which discusses conditions under which two games with identical best response correspondences must have payoffs that can be obtained from each other by positive affine transformations. The proofs of the two results share similar mathematical ideas.

Now consider the same question profile X in a different decision problem where the subject can also receive a smaller reward for a "close" guess. Let $\Theta = \{1, \ldots, n\}$ and

$$\tilde{u}(a; \theta) = r_\theta \mathbb{1}\{a = \theta\} + \frac{r_\theta}{2} \mathbb{1}\{|a - \theta| = 1\}$$

where $\min_\theta r_\theta > \max_\theta r_\theta / 2 > 0$. The adjacency graph for this problem is again complete. In this case, however, X is not aligned with $\tilde{u}$. Since the linear independence condition in Theorem 2 holds, X is not incentivizable.

To explain the main ideas in the proof of Theorem 2, we first need some additional terminology. An *adjacency cycle* is a tuple $C = (a_0, \ldots, a_n)$ such that $a_n = a_0$ and actions a_i and a_{i+1} are adjacent for each $i = 0, \ldots, n-1$. We say that n is the *length* of the cycle C. We abuse notation slightly by writing $a \in C$ to mean that $a = a_i$ for some i.

An adjacency cycle $C = (a_0, \ldots, a_n)$ is *internally independent* if the set of vectors $\left\{\Delta_{a_0}^{a_i}\right\}_{i=1,\ldots,n-1}$ is linearly independent. Let V_C be the linear space spanned by this set of vectors. The cycle C is internally independent if and only if $\dim V_C = n - 1$. Notice that the space V_C and the linear independence of $\left\{\Delta_{a_0}^{a_i}\right\}_{i=1,\ldots,n-1}$ depend only on C itself, and not on which action in the cycle is labeled as a_0.

As described above, for each action a_0 in an adjacency cycle, applying the Adjacency Lemma iteratively along pairs $\{a, b\}$ of adjacent actions in the cycle gives rise to an expression for $\bar{X}(a_0)$ as a linear combination of Δ_a^b terms. If this cycle is internally independent, then either $\bar{X}(a_0) \in V_C \setminus \{0\}$ or all of the coefficients in the linear combination must be equal to 0. In the latter case, iteratively applying the Adjacency Lemma again gives a linear relationship between any $\bar{X}(c)$ (for c in the cycle) and $\bar{X}(a_0)$ together with Δ_a^b terms. This relationship can be simplified using the fact that the coefficients in the previous linear combination are zero to show that X is aligned with u on C, leading to the following lemma.

Lemma 2. *Let C be an internally independent adjacency cycle. If X is incentivizable, then either X is aligned with u on C, or $\bar{X}(a) \in V_C \setminus \{0\}$ for all $a \in C$.*

While Lemma 2 identifies conditions that an incentivizable question profile must satisfy on a given internally independent cycle, it says nothing about more complicated adjacency graphs. The following lemma shows how alignment with u on multiple cycles or other subsets of actions can, under mild conditions, be combined to obtain alignment with u on their union.

Lemma 3. *Suppose X is aligned with u on sets of actions B and D. If there exist actions $a_0, a_1 \in B \cap D$ such that $\bar{X}(a_0)$ and $\bar{X}(a_1)$ are not collinear, then X is aligned with u on $B \cup D$.*

Alignment with u on B gives rise, for each action in B, to a system of equations that the parameters relating $\bar{X}$ and $\bar{u}$ must satisfy; naturally, the same is true for D. In the proof of this lemma, we show that if B and D share two actions for which the values of the questions are linearly independent, then the corresponding systems of equations must share the same solution for B as for D.

The main idea behind Theorem 2 is that, given any cycle C through an action a, Lemma 2 implies that any incentivizable X is either aligned with u on C or lies in $V_C \setminus \{0\}$. The linear independence condition can be used to eliminate the latter possibility by considering multiple cycles through a; more precisely, there must be *some* cycle C on which X is aligned with u, as the intersection of the sets $V_{C'}$ across cycles C' containing a is $\{0\}$. Varying a and applying Lemma 3 leads to alignment with u on the entire action set.

In Appendix D, we identify a more general class of problems in which alignment with u is both necessary and sufficient for incentivizability. This more general result applies when there is a rich enough structure of internally independent cycles in the adjacency graph, which can be the case even when the graph is not complete.

8. PRODUCT PROBLEMS

In many experiments, subjects perform a sequence of tasks. These tasks may be identical or they may differ. The subject's payoff is a sum or weighted average of the payoffs in the various tasks, as in the common experimental design in which one task is randomly selected for payment (see Charness, Gneezy, and Halladay (2016) and Azrieli, Chambers, and Healy (2018) and the references therein). In such cases, the researcher may be interested in eliciting beliefs related to the entire sequence of actions chosen by the subject. For example, a subject may solve a test with multiple problems—with their payoff being equal to their score—and the researcher may want to ask the subject what he believes about his overall performance on the test. Alternatively, to gauge the impact of learning across repetitions of the same task, the researcher may want to elicit subjects' beliefs about their change in performance between the beginning and the end of the experiment.

To formalize this idea, we define a *product problem* (Θ, A, u) to be a decision problem in which $\Theta = \times_i \Theta_i$, $A = \times_i A_i$, and $u(a; \theta) = \sum_i u_i(a_i, \theta_i)$ for some sets

$(\Theta_1, \ldots, \Theta_I)$ and $(A_1, \ldots, A_I)$, and some functions $u_i : A_i \times \Theta_i \longrightarrow \mathbb{R}$.[15] As noted above, the additive separability of u captures commonly used incentives in which one choice is randomly selected for payment. We refer to each (Θ_i, A_i, u_i) as a *task*. We write a_{-i} for a profile $(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_I)$ and $a_i a_{-i}$ for the profile whose ith coordinate is a_i and remaining coordinates are given by a_{-i}.

Example 4. The decision problem is a test consisting of $I \geq 3$ multiple choice questions. The state space and action space are given by $\Theta = A = \Omega^I$, where Ω is a finite set containing at least two elements describing the possible answers to any given question. Coordinate i corresponds to the ith question: θ_i is the correct answer to question i and a_i is the subject's answer to question i. The payoff in the decision problem is the score on the test: $u(a; \theta) = \sum_{i=1}^{I} \mathbb{1}\{\theta_i = a_i\}$. The subject has a belief $p \in \Delta(\Theta)$. The u-optimal choice of action in each task i is the most likely state according to the marginal distribution of p over Θ_i.
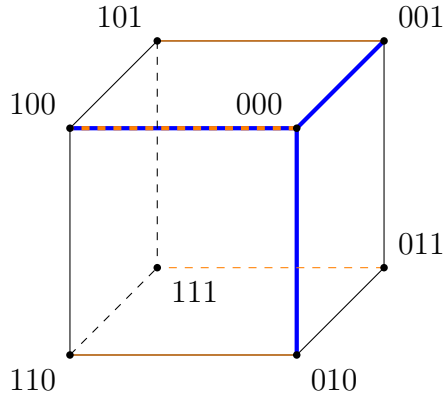
Note that we make no assumptions about the correlation among states across tasks: the subject can hold any belief about the joint distribution of $(\theta_1, \ldots, \theta_I)$. In particular, the subject need not view the states as independent, nor must there be a fixed state across tasks.

Product problems share a structure that distinguishes them from the problems we have analyzed so far. Two actions $a, b \in A$ are adjacent only if they differ in exactly one task, that is, if there is some i such that $a_i \neq b_i$ and $a_{-i} = b_{-i}$. Conversely, if a_i and b_i are adjacent in task i, then the product actions $a_i a_{-i}$ and $b_i a_{-i}$ are adjacent for all a_{-i}. The adjacency graph of the product problem is generally neither complete (even if the adjacency graph in each task is complete) nor a tree.

Figure 2 depicts the adjacency graph in Example 4 when $\Omega = \{0, 1\}$ and $I = 3$. Note that while the collection of payoff difference vectors associated with edges exiting a single node—such as the blue edges Δ_{000}^{100}, Δ_{000}^{010}, and Δ_{000}^{001}—are linearly independent, no cycle is internally independent because parallel edges correspond to identical payoff difference vectors. Thus, for instance, the orange edges correspond to $\Delta_{000}^{100} = \Delta_{010}^{110} = \Delta_{001}^{101} = \Delta_{011}^{111}$.

We say that question profile X depends on task i *trivially* if, for each a_{-i}, the vectors $\bar{X}(a_i a_{-i})$ are collinear for all a_i. Trivial dependence on task i means that

---

[15]Theorem 3 below generalizes to decision problems with payoffs of the form $u(a; \theta) = \sum_i u_i(a_i, \theta)$, i.e., where the payoff within each task can depend on the entire state of the world θ, under the assumption that the collection of subspaces $E_i = \text{span}\{\Delta_{a_i}^{b_i} : a_i, b_i \in A_i\}$ is linearly independent.

FIGURE 2. Adjacency graph for Example 4 with $\Omega = \{0, 1\}$ and $I = 3$.

changing a_i entails replacing $\mathsf{X}(a)$ with an equivalent question. The following result provides necessary and sufficient conditions for incentivizability in product problems.

Theorem 3. *Let $(\Theta_i, A_i, u_i)_{i=1}^{I}$ be a product problem with $I \geq 3$. Suppose that for each i, either A_i contains only two actions, or the adjacency graph for problem (Θ_i, A_i, u_i) is complete and the vectors $\left\{\Delta_{a_i}^{b_i}, \Delta_{a_i}^{c_i}\right\}$ are linearly independent for all distinct $a_i, b_i, c_i \in A_i$. Suppose in addition that there are at least three tasks on which X does not depend trivially. Then X is incentivizable if and only if there exist $v(a), \kappa(a) \in \mathbb{R}$ with $v(a) \neq 0$ for each $a \in A$, $\tau_i \in \mathbb{R}$ for each i, and $\mathsf{d} \in \mathbb{R}^{\Theta}$ such that*

$$X(a; \theta) = \kappa(a) + v(a) \left(d(\theta) + \sum_i \tau_i u_i(a_i, \theta_i) \right) \tag{6}$$

for all a and θ.

Along the same lines as Theorem 2, we assume that each task has a complete adjacency graph (which holds vacuously when the action set is binary). However, relative to Theorem 2, the other requirements for each task are significantly weaker: there are no restrictions on the number of actions, and we require linear independence only of pairs of payoff difference vectors instead of triples.

The characterization of incentivizable question profiles in (6) is more permissive than that in Theorem 2. If τ_i is constant across i, the expression in (6) implies that X is aligned with u. However, in contrast to the case of a single problem with complete adjacency, there are many questions not aligned with u that are also incentivizable (namely, those for which τ_i varies across i). The additional freedom in the product problem results from a smaller number of cycles and a larger number of linear dependencies in the payoffs.

We refer to any X that satisfies the conditions of Theorem 3 as *weighted aligned.* The following example illustrates the added flexibility of weighted alignment relative to alignment, as well as its restrictiveness.

Example 4 (continued). It is straightforward to verify that in the product problem of Example 4, if $|\Omega| > 2$, the adjacency graph for each problem i is complete and all pairs $\left\{\Delta_{a_i}^{b_i}, \Delta_{a_i}^{c_i}\right\}$ are linearly independent. Therefore, Theorem 3 applies to any question profile that depends nontrivially on at least three tasks.

Suppose the test consists of two parts: the first part comprises questions 1 through I_1, while the second part comprises questions $I_1 + 1$ through $I = I_1 + I_2$. Consider the question that asks the subject about the expected improvement in his average score from the first part of the test to the second:

$$X(a; \theta) = \frac{1}{I_2} \sum_{i > I_1} \mathbb{1}(a_i = \theta_i) - \frac{1}{I_1} \sum_{i \le I_1} \mathbb{1}(a_i = \theta_i).$$

This question profile is weighted aligned, and is therefore incentivizable by Theorem 3. At first glance, this result may be surprising as X seems to create opposing incentives in the two parts of the test. Nevertheless, X can be incentivized using a simple modification of the BDM elicitation mechanism from Proposition 1 that involves adding the payoff from the decision problem:

$$V(r, a, \theta) = \int_0^r X(a; \theta) dx + \int_r^1 x dx - \frac{1}{2} + \sum_i u_i(a_i, \theta_i) \tag{7}$$

$$= \sum_i \left[1 - \mathbb{1}\{i \le I_1\} \frac{r}{I_1} + \mathbb{1}\{i > I_1\} \frac{r}{I_2} \right] u_i(a_i, \theta_i) - \frac{r^2}{2}.$$

The two integral terms provide incentives for truthful reporting of $\mathbb{E}_p X(a; \cdot)$, as in the standard BDM mechanism. For the optimal action a, we can think of simultaneously choosing r and a to maximize V. For each i, because the corresponding coefficient on u_i in the square brackets is positive for every $r \in [0, 1]$, the optimal action a_i is the one that maximizes the expectation of $u_i(a_i, \theta_i)$. Therefore, the overall payoff under V provides the correct incentives for the action a.

For an example of a question that is not weighted aligned, and hence not incentivizable, consider elicitation of the probability the subject assigns to receiving a score of

at least z, where $z \in \{1, \ldots, I\}$.[16] This question corresponds to

$$Y(a; \theta) = \begin{cases} 1 & \text{if } \sum_{i=1}^{I} \mathbb{1}\{\theta_i = a_i\} \geq z \\ 0 & \text{otherwise.} \end{cases}$$

We claim that Y is not incentivizable. Since there is no task on which Y depends trivially, Theorem 3 applies. Thus it suffices to show that Y is not of the form specified in (6). Suppose for contradiction that it is. The idea is that, for a given action, a change in the state that affects the utility but not the value of Y implies restrictions on the parameters in (6) that are violated for actions where the same change in the state changes the value of Y. To make this precise, suppose, in addition, that $z < I$; it is straightforward to adapt the argument for the case of $z = I$. Given any a and i, let $\theta = a$ and let θ' be such that $\theta'_i \neq \theta_i$ and $\theta'_j = \theta_j$ for all $j \neq i$. Thus a gets a perfect score of I if the state is θ and a score of $I - 1 \geq z$ if the state is θ'. In particular, $Y(a; \theta) = Y(a; \theta') = 1$. Equating the right-hand sides of (6) for $Y(a; \theta)$ and $Y(a; \theta')$, since $v(a) \neq 0$, it follows that $d(\theta) + \sum_j \tau_j = d(\theta') + \sum_{j \neq i} \tau_j$, and hence $d(\theta) + \tau_i = d(\theta')$. Repeating the argument for $a' = \theta'$ with the roles of θ and θ' reversed, we obtain $d(\theta') + \tau_i = d(\theta)$. Therefore, $\tau_i = 0$ and $d(\theta) = d(\theta')$. Now consider an action a'' that agrees with θ on exactly z coordinates, including coordinate i. Then a'' agrees with θ' on $z - 1$ coordinates, and we have $Y(a''; \theta) = 1 \neq 0 = Y(a''; \theta')$. From (6), since $\tau_i = 0$, it must be that $d(\theta) \neq d(\theta')$, a contradiction.

Intuitively, one may think that using an elicitation method (such as BDM) whose value is increasing in $\mathbb{E}_p[Y(a; \cdot)]$ would suffice for incentivizability, as then the subject can increase her payment at the belief elicitation stage by "doing well" on the test. The above argument shows that this intuition is incorrect. The problem lies in the definition of doing well: maximizing the expected score in the test is not generally the same thing as maximizing the likelihood of that score being above a fixed threshold.

8.1. **Proof idea.** The proof of Theorem 3 can be found in Appendix E. Here, we describe the key ideas. For simplicity, we focus on Example 4 with actions $A_i = \{0_i, 1_i\}$.

The proof that equation (6) is sufficient for incentivizability is relatively straightforward: the argument directly extends the construction from equation (7).

For the remainder of this section, we assume that X is incentivizable. By the Adjacency Lemma, for any pair of adjacent actions a and b, there are coefficients

---

[16]This question is similar to that of Möbius et al. (2022), who elicit the subject's belief that their score on an incentivized IQ test is above the median among the participants.

$x(a, b) \neq 0$ and $y(a, b)$ such that $\bar{\mathsf{X}}(a) = x(a, b)\bar{\mathsf{X}}(b) + y(a, b)\Delta_a^b$. For the purpose of this discussion, we assume that the coefficients $x(a, b)$ and $y(a, b)$ are uniquely defined. Note also that, for any such pair, there is a single task i on which the actions a_i and b_i differ, and $\Delta_a^b = \Delta_{a_i}^{b_i}$, where $\Delta_{a_i}^{b_i}$ refers to the payoff difference $\bar{u}_i(b_i) - \bar{u}_i(a_i)$. We say that i is the *relevant* task for the pair (a, b).

The key step in the proof is to show that each cycle $(a^0, \ldots, a^n = a^0)$ of adjacent actions is *exact*, that is, that $x(a^0, a^1)x(a^1, a^2) \cdots x(a^{n-1}, a^n) = 1$. We first explain the connection between exactness of all cycles and (6), and then explain why all cycles are exact.

The exactness of all cycles implies that we can define $v(a) = x(0, a^1) \cdots x(a^{n-1}, a^n)$ for any path $a^0 = 0, \ldots, a^n = a$ of adjacent actions, and the definition does not depend on the chosen path. In particular, for any two adjacent actions a and b, $v(a) = x(b, a)v(b)$. Letting $\bar{\mathsf{X}}^*(a) = \frac{1}{v(a)}\bar{\mathsf{X}}(a)$, a repeated application of the Adjacency Lemma implies that,

$$\bar{\mathsf{X}}^*(a) = \bar{\mathsf{X}}^*(0) + \sum_{m<n} y^*(a^m, a^{m+1})\Delta_{a^m}^{a^{m+1}} \tag{8}$$

for some (arbitrary) path $a^0 = 0, \ldots, a^n = a$.

Consider paths of length $n = |\{i : a_i = 1_i\}|$ from 0 to a. Along such paths, each task in the product problem is relevant at most once. Let i_m be the relevant task for pair (a^m, a^{m+1}). Then, $\Delta_{a^m}^{a^{m+1}} = \Delta_{0_{i_m}}^{1_{i_m}}$. Because the vectors $\Delta_{0_i}^{1_i}$ are linearly independent across tasks, by equating the right-hand side of (8) across all length-n paths, we conclude that the coefficients y^* depend only on the relevant task for the transition; thus we can define $y_{i_m} = y^*(a^m, a^{m+1})$ and obtain

$$\bar{\mathsf{X}}^*(a) = \bar{\mathsf{X}}^*(0) + \sum_{i \text{ s.t. } a_i=1} y_i\Delta_{0_i}^{1_i} = \bar{\mathsf{X}}^*(0) - \sum_i y_i\bar{u}_i(0_i) + \sum_i y_i\bar{u}_i(a_i).$$

Taking $d(\theta) = \bar{\mathsf{X}}^*(0) - \sum_i y_i\bar{u}_i(0_i)$ gives an expression of the form in (6).

We now return to the question of why all cycles are exact. As the first step, consider the four-action cycle corresponding to the front face in Figure 2. A repeated application of the Adjacency Lemma along the cycle leads to

$$\bar{\mathsf{X}}(000) = x(000, 100)x(100, 110)x(110, 010)x(010, 000)\bar{\mathsf{X}}(000) + s_1\Delta_{0_1}^{1_1} + s_2\Delta_{0_2}^{1_2}$$

for some $s_1, s_2 \in \mathbb{R}$. If $\bar{\mathsf{X}}(000)$ is not in the subspace spanned by the vectors $\Delta_{0_1}^{1_1}$ and $\Delta_{0_2}^{1_2}$, then the product of the x coefficients on the right-hand side must be equal to 1, meaning that the cycle is exact. Otherwise, by analyzing a number of cases, one can show that the cycles corresponding to the other five faces in Figure 2 are all exact.

These five cycles can be combined in such a way that all edges not belonging to the front face "cancel out," thereby implying exactness of the original cycle. In general, we show that all similar 4-cycles are exact.

In the second step, we show how to decompose arbitrary cycles into 4-cycles in such a way that the exactness of the original cycle follows from the exactness of cycles in the decomposition.

These arguments are not specific to binary actions. Extending beyond binary actions requires more care and conditions on linear independence of payoffs within tasks. In the proof of exactness, aside from the 4-cycles described above, we also use 3-cycles of adjacent actions with the same relevant task. We leave the details to the formal proof in the appendix.

9. Discussion

9.1. **Two or three states.** As noted in Section 5, the Adjacency Lemma has limited bite if there are only two or three states due to the dimensionality of its restrictions. With two states, the set of beliefs at which the subject is indifferent between two actions is a single point. With three states, the indifference set intersected with the set of beliefs along which there is a constant answer to a particular question is generically a single point. In both cases, the thesis of the Adjacency Lemma is trivially satisfied.

If there are only two states, the elicitation problem becomes trivial as the subject's belief about the state can be elicited independent of the action, which is sufficient for the researcher to determine the expectation of any question $\mathsf{X}(a)$.

With three states, eliciting the entire belief may still be a practical option as it requires asking for only two probabilities. If the researcher wants to ask for only one number, although the Adjacency Lemma is vacuous, looking at adjacencies can nonetheless be useful for understanding incentivizability. Suppose the overall payoff is a weighted sum of the payoff from the decision problem and that from a scoring rule applied at the belief elicitation stage. Suppose moreover that the scoring rule depends only on the reported belief and the realized value of $X(a; \theta)$, and not on θ directly.[17] At any belief at which the subject is indifferent between two actions, the value of truthful reporting at the belief elicitation stage must be equal following

[17] This is a natural restriction that we do not impose in our model as doing so would have no effect on our results. When there are three states, we do not know whether this restriction has any bite; we expect that the convexity of the value function would place restrictions on incentivizable questions even if the value can depend on θ directly.
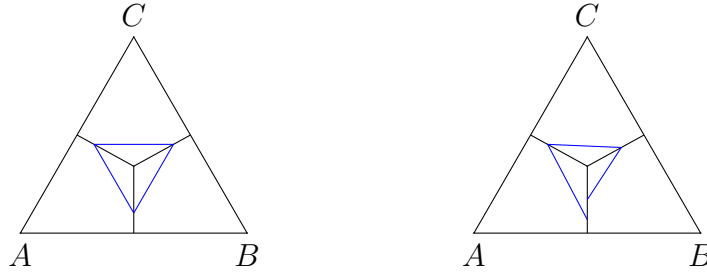
FIGURE 3. Problems with three states. Each triangle depicts the simplex of beliefs. The black line segments illustrate the partition of the simplex according to which action is u-optimal. Blue line segments represent sets of beliefs on which $\mathsf{X}(a^*)$ is constant for the u-optimal action a^*. In the left triangle, because the blue lines form a triangle, adjacency considerations do not rule out incentivizability of X. In the right triangle, X is not incentivizable if the payment for elicitation depends only on the value of X and the reported belief.

these two actions. Depending on the structure of the decision problem, following these constant values along cycles of adjacent actions can imply restrictions on X; see Figure 3.

9.2. **Independent questions.** Our necessary conditions make use of independence assumptions on the payoffs in the decision problem. Similar results can be obtained if one replaces these assumptions with assumptions about independence of $\bar{\mathsf{X}}(a)$ across actions. For example, along the lines of Lemma 2, if X is incentivizable and the set of questions $\bar{\mathsf{X}}(a)$ is linearly independent for actions a in some cycle C of adjacencies, then one can show that X must be aligned with u on C. Lemma 3 can then be used to obtain necessary conditions on the full set of actions.

9.3. **Non-affine questions.** We have restricted attention to eliciting beliefs about the expectation of some function $\mathsf{X}(a)$. Lambert (2019) studies elicitation of "properties" of beliefs, where a property corresponds to a partition of the simplex. He characterizes which properties are incentivizable, i.e., for which ones there exists a scoring rule incentivizing truthful reporting when the subject is asked only about the property associated with his belief. A question $\mathsf{X}(a)$ in our framework corresponds to a property that partitions the simplex into parallel hyperplanes (unless the question is trivial, in which case there is a single property for the entire simplex). This formulation captures many properties of interest and ensures incentivizability both in Lambert's sense and, if the question is independent of the action choice, in our

sense. There are, however, properties—such as the median of some X—which may be of interest that are incentivizable in Lambert's context but to which our results do not apply. Nonetheless, we expect our general approach of focusing on adjacencies between actions to be useful for such non-affine questions.

9.4. **Observability.** We assume the researcher knows the subject's utility function and observes actions and states. These assumptions are typically satisfied in lab experiments, but could be violated in field experiments. A lack of observability introduces additional issues that we view as orthogonal those we study. Our negative results describe the limits of nondistortionary elicitation; if the researcher does not observe all aspects of the experiment, elicitation becomes even more difficult.

Interestingly, willingness-to-pay (or expected utility) is always incentivizable in our framework, and is also often the only question that can be elicited under incomplete observability (see Chassang, Padró I Miquel, and Snowberg, 2012). In that sense, both issues may lead to similar prescriptions regarding what can be elicited.

9.5. **Endogenous beliefs.** Another important assumption of our framework is that the subject has a fixed belief that is not affected by his choice of action. Consider, for example, the moral hazard issue described in Chambers and Lambert (2021) in which a student who is asked about the probability of passing a test can influence that probability through their behavior. If the state is defined as passing or failing, the student's belief is endogenous.

In many cases, a problem with endogenous beliefs can be converted into one with fixed beliefs by redefining variables. In the above example, the state could instead be the solution to the test. If the outcome of the test is a deterministic function of the state and the student's behavior, the likelihood of passing corresponds to a question profile X in our model. More generally, one can redefine the state as a function from the set of actions into probability distributions over outcomes. If this reformulation satisfies the observability assumptions of our model, then our results apply.

9.6. **Multi-dimensional questions.** In our model, we assume that the researcher can only ask a single question. Here, we show how our methods can be extended to multiple questions. For simplicity, we focus on the case of two questions; the logic extends directly to more than two questions.

We say that question profiles $X, Y : A \longrightarrow \mathbb{R}^\Theta$ are *jointly incentivizable* if there exists an elicitation method $V : \mathbb{R}^2 \times A \times \Theta \longrightarrow [0, 1]$ such that, for every $p \in \Delta(\Theta)$,

$$\arg\max_{a,r,s} \mathbb{E}_p V(r, s, a, \theta) = \left\{ (\mathbb{E}_p X(a; \theta), \mathbb{E}_p Y(a; \theta), a) : a \in \arg\max_{b \in A} \mathbb{E}_p u(b; \cdot) \right\}.$$

If X and Y are both aligned with u, it is straightforward to extend Proposition 1 to show that they are jointly incentivizable.

For necessary conditions, our key result—Lemma 1—extends as follows.

Lemma 4. *Suppose X and Y are jointly incentivizable. If actions a and b are adjacent, then there are ρ_X, ρ_Y and σ_x^y for $x, y = X, Y$, not all equal to 0, such that*

$$\bar{X}(b) = \rho_X (\bar{u}(b) - \bar{u}(a)) + \sigma_X^X \bar{X}(a) + \sigma_X^Y \bar{Y}(a)$$

$$and \qquad \bar{Y}(b) = \rho_Y (\bar{u}(b) - \bar{u}(a)) + \sigma_Y^X \bar{X}(a) + \sigma_Y^Y \bar{Y}(a).$$

If $\bar{X}(a)$ or $\bar{X}(b)$ is collinear with $\bar{u}(b) - \bar{u}(a)$, then we can take $\rho \neq 0$.

The proof, which we omit, follows the same reasoning as that of Lemma 1. We leave to future research the details of how to use this lemma to identify precise conditions for joint incentivizability.

APPENDIX A. PROOFS FOR SECTION 4

Lemma 5. *For any $\mathsf{d} \in \mathbb{R}^\Theta$, $X(a) = \mathsf{u}(a) + \mathsf{d}$ and $X(a) = \mathsf{d}$ are incentivizable.*

Proof. Let $L, M \in \mathbb{R}$ be such that, for all a and θ, $L < u(a; \theta) + d(\theta) < M$ and $L < d(\theta) < M$. If $X(a) = \mathsf{u}(a) + \mathsf{d}$, let

$$V(r, a, \theta) = \int_L^r X(a; \theta) \, dx + \int_r^M x \, dx - \frac{M^2}{2} = (u(a; \theta) + d(\theta))(r - L) - \frac{1}{2} r^2,$$

where $d(\theta)$ is the θ-coordinate of d. For this V, simple calculations show that the optimal choice of r given a is $\mathbb{E}_p X(a; \theta)$. Since the optimal r is greater than L, the V-optimal choice of action a is the same as the u-optimal one. A similar argument applies if $X(a) = \mathsf{d}$, in which case we let $V(r, a, \theta) = u(a; \theta) + d(\theta)(r - L) - \frac{1}{2} r^2$. $\quad\square$

Proof of Proposition 1. Suppose V incentivizes a question profile X. Let $Y(a) = \gamma(a) X(a) + \kappa(a) \mathbf{1}$ for some $\gamma, \kappa : A \longrightarrow \mathbb{R}$. Letting $W(r, a, \theta) = V\left(\frac{1}{\gamma(a)}(r - \kappa(a)), a, \theta\right)$, it is straightforward to verify that W incentivizes Y. The result now follows from Lemma 5. $\quad\square$

APPENDIX B. PROOFS FOR SECTION 5

Proof of Lemma 1. We first prove the latter formulation of the result (i.e., that $\bar{X}(b) = \rho\Delta_a^b + \sigma\bar{X}(a)$), and then prove the equivalence between the two formulations.

Let V be an elicitation method that incentivizes X. Consider two beliefs p_0 and p_1 such that (i) actions a and b are both u-optimal, i.e., $a, b \in \arg\max_{a'} \mathbb{E}_{p_k}[u(a'; \cdot)]$ for $k = 0, 1$, and (ii) the question $X(a)$ attains the same value at p_0 and p_1, i.e., $\mathbb{E}_{p_0}[X(a; \cdot)] = \mathbb{E}_{p_1}[X(a; \cdot)]$. Letting $r = \mathbb{E}_{p_0}[X(a; \cdot)]$, it follows that (a, r) is optimal given V at all $p_\alpha = \alpha p_1 + (1 - \alpha) p_0$ for $\alpha \in [0, 1]$. The value of information $V^*(p_\alpha) = \mathbb{E}_{p_\alpha}[V(r, a, \cdot)]$ is therefore affine in α. Since b is also a u-optimal action at each p_α, the optimal expected payoff $\max_s \mathbb{E}_{p_\alpha}[V(s, b, \cdot)]$ must be affine in α as well. Therefore, as explained in Section 5, there exists some r' such that, for each α, $r' \in \arg\max_s \mathbb{E}_{p_\alpha}[V(s, b, \cdot)]$. In particular, $\mathbb{E}_{p_k}[X(b; \cdot)] = r'$ for each $k = 0, 1$.

The adjacency of actions a and b implies that we can find an interior p_0 for which the set of u-optimal actions is $\{a, b\}$. Consider the linear subspace

$$D = \{d \in \mathbb{R}^\Theta : d \cdot 1 = d \cdot (u(a) - u(b)) = d \cdot X(a) = 0\}$$
$$= (\text{span}\{1, u(a) - u(b), X(a)\})^\perp.$$

For each $d \in D$, there is a sufficiently small $\varepsilon > 0$ for which $p_1 = p_0 + \varepsilon d$ (i) is a well-defined belief, (ii) a and b are both u-optimal at p_1, and (iii) $X(a)$ attains the same value at p_0 as at p_1; each of these properties follows from one of the orthogonality conditions defining D. In particular, p_0 and p_1 satisfy properties (i) and (ii) from the previous paragraph. As a result, the above argument applies and shows that $\varepsilon d = p_1 - p_0 \perp X(b)$. Therefore, $X(b) \perp D$.

By a standard linear algebra argument, it follows that

$$X(b) \in D^\perp = \text{span}(1, u(a) - u(b), X(a)).$$

Noting that, for any vector v, $\bar{v}$ differs from v by a scalar multiple of 1, and that $\bar{v} \perp 1$, we obtain $\bar{X}(b) = \rho(\bar{u}(a) - \bar{u}(b)) + \sigma\bar{X}(a)$ for some $\rho, \sigma \in \mathbb{R}$.

If $\bar{X}(b)$ is not collinear with $\bar{u}(b) - \bar{u}(a)$, then we must have $\sigma \neq 0$, as needed. Otherwise, switching the roles of a and b in the preceding argument yields that $\bar{X}(a)$ is also collinear with $\bar{u}(b) - \bar{u}(a)$, and therefore one can take ρ and σ to be nonzero.

We now show that the above conclusion is equivalent to alignment on pairs of adjacent actions. Note first that X is nontrivially aligned with u on a set of actions B if and only if $\bar{X}(a) \equiv_B \gamma(a)\bar{u}(a) + d$ for some $\gamma(a) \in \mathbb{R}$ and $d \in \mathbb{R}^\Theta$, and is trivially aligned with u on B if and only if $\bar{X}(a) \equiv_B \gamma(a)d$ for some $\gamma(a) \in \mathbb{R}$ and $d \in \mathbb{R}^\Theta$. In

either case, eliminating d leads to an equation of the form $\bar{\mathsf{X}}(b) = \rho\Delta_a^b + \sigma\bar{\mathsf{X}}(a)$ with $\sigma \neq 0$.

For the converse, if $\rho = 0$, then X is trivially aligned with u on $\{a, b\}$. If $\rho \neq 0$, then the alignment is nontrivial, with $\mathsf{d} = \frac{1}{\rho}\bar{\mathsf{X}}(b) - \bar{\mathsf{u}}(b) = \frac{\sigma}{\rho}\bar{\mathsf{X}}(a) - \bar{\mathsf{u}}(a)$, $\gamma(a) = \rho/\sigma$, and $\gamma(b) = \rho$. $\qquad\square$

APPENDIX C. PROOFS FOR SECTION 6

Lemma 6. *Let a be a splitting action and B_0 and B_1 be sets not containing a such that $\{B_0 \cup \{a\}, B_1 \cup \{a\}\}$ is a splitting collection. Then, for any pair of beliefs p_0 and p_1 such that, for each i, some action $b_i \in B_i$ is u-optimal at p_i, there is a convex combination $p = \alpha p_1 + (1 - \alpha) p_0$ such that a is u-optimal at p.*

Proof. Suppose not. Then all actions that are u-optimal at convex combinations of p_0 and p_1 must be either from B_0 or B_1. Hence, for some such convex combination p', the set of u-optimal actions contains at least one element from each of B_0 and B_1. Then in any neighbourhood of p', there must be a belief at which there is exactly one member of each B_i that is u-optimal, contradicting the fact that no action in B_0 is adjacent to any action in B_1. $\qquad\square$

Proof of Proposition 2. We first derive explicit formulas for an elicitation method that incentivizes a question X aligned with u. When X is nontrivially aligned with u, using the BDM construction from the proof of Proposition 1 gives

$$V^{BDM}(r, a, \theta) = (u(a; \theta) + d(\theta)) \left(\frac{1}{\gamma(a)}(r - \kappa(a)) - L\right) - \frac{1}{2}\left(\frac{1}{\gamma(a)}(r - \kappa(a))\right)^2 \tag{9}$$

$$= \frac{1}{\gamma(a)^2}\left[(X(a; \theta) - \kappa(a))(r - \kappa(a) - \gamma(a)L) - \frac{1}{2}(r - \kappa(a))^2\right]$$

$$= \frac{1}{\gamma(a)^2}\left(X(a; \theta) - \frac{1}{2}r - \frac{1}{2}\kappa(a)\right)(r - \kappa(a)) - \frac{1}{\gamma(a)}(X(a; \theta) - \kappa(a))L.$$

Similarly, if X is trivially aligned with u, adding $u(a; \theta)$ to either of the last two lines gives V^{BDM}.

When X is nontrivially aligned with u, the expected payoff of a subject who chooses action a and then chooses r optimally is equal to

$$\max_r \left[\mathbb{E}_p \left[u\left(a;\cdot\right) + d\left(\cdot\right) \right] \left(\frac{1}{\gamma(a)} \left(r - \kappa(a)\right) - L \right) - \frac{1}{2} \left(\frac{1}{\gamma(a)} \left(r - \kappa(a)\right) \right)^2 \right] \qquad (10)$$

$$= \max_x \left[\mathbb{E}_p \left[u\left(a;\cdot\right) + d\left(\cdot\right) \right] \left(x - L\right) - \frac{1}{2}x^2 \right]$$

$$= \frac{1}{2} \left(\mathbb{E}_p \left[u\left(a;\cdot\right) + d\left(\cdot\right) \right] \right)^2 - L\mathbb{E}_p \left[u\left(a;\cdot\right) + d\left(\cdot\right) \right].$$

Similarly, the expected payoff is equal to $\mathbb{E}_p[u(a;\cdot)] + \left(\mathbb{E}_p[d(\cdot)]\right)^2/2 - L\mathbb{E}_p[d(\cdot)]$ if X is trivially aligned with u.

To keep the notation simple, we present the argument only for the case in which the splitting collection contains two elements, A_0 and A_1, with splitting action a_0. Extending the argument to the general case is straightforward.

We construct an elicitation method on each A_i, then show that they agree on a_0 and therefore give rise to well-defined elicitation method on the full action set, A.

First suppose X is nontrivially aligned with u on each A_i. For each $i = 0, 1$, let $(\gamma_i, \kappa_i, \mathsf{d}_i)$ be such that $\mathsf{X}(a) \equiv_{A_i} \gamma_i(a)(\mathsf{u}(a) + \mathsf{d}) + \kappa(a)\mathbf{1}$. Define the elicitation methods

$$V_i\left(r, a, \theta; w_i, \omega_i\right) = \left(\frac{\gamma_i(a_0)}{\gamma_i(a)} \right)^2 \left(X\left(a;\theta\right) - \frac{1}{2}r - \frac{1}{2}\kappa_i(a) \right) \left(r - \kappa_i(a)\right)$$

$$+ w_i\left(\theta\right) - \omega_i - \frac{\gamma_i(a_0)^2}{\gamma_i(a)}\left(X(a;\theta) - \kappa_i(a)\right)L, \quad (11)$$

where $w_i \in \mathbb{R}^\Theta$ and $\omega_i \in \mathbb{R}$. Note that this expression differs from the expression for V^{BDM} in (9) only by multiplication by a positive constant and addition of a function that depends only on the state. As neither of these changes affects the optimal choices of a and r, it follows from the argument in the proof of Proposition 1 that V_i incentivizes X on A_i (i.e., in the decision problem with actions restricted to A_i).

Notice that

$$V_1\left(r, a_0, \theta; w_1, \omega_1\right) - V_0\left(r, a_0, \theta; w_0, \omega_0\right)$$

$$= \left(X\left(a_0; \theta\right) - \frac{1}{2}r - \frac{1}{2}\kappa_1(a_0)\right)(r - \kappa_1(a_0)) - \left(X\left(a_0; \theta\right) - \frac{1}{2}r - \frac{1}{2}\kappa_0(a_0)\right)(r - \kappa_0(a_0))$$

$$+ (w_1(\theta) - w_0(\theta)) - (\omega_1 - \omega_0) - \gamma_1(a_0)(X(a_0; \theta) - \kappa_1(a_0))L + \gamma_0(a_0)(X(a_0; \theta) - \kappa_0(a_0))L$$

$$= X\left(a_0; \theta\right)\left(\kappa_0(a_0) - \kappa_1(a_0) + L\gamma_0(a_0) - L\gamma_1(a_0)\right) + \frac{1}{2}\left((\kappa_1(a_0))^2 - (\kappa_0(a_0))^2\right)$$

$$+ (w_1(\theta) - w_0(\theta)) - (\omega_1 - \omega_0) + (\gamma_1(a_0)\kappa_1(a_0) - \gamma_0(a_0)\kappa_0(a_0))L.$$

Given any w_0 and ω_0, let

$$w_1(\theta) = w_0(\theta) - X\left(a_0; \theta\right)\left(\kappa_0(a_0) - \kappa_1(a_0) + L\gamma_0(a_0) - L\gamma_1(a_0)\right)$$

for each θ, and

$$\omega_1 = \frac{1}{2}\left(\kappa_1(a_0)^2 - \kappa_0(a_0)^2\right) + \omega_0 + (\gamma_1(a_0)\kappa_1(a_0) - \gamma_0(a_0)\kappa_0(a_0))L.$$

Then $V_1\left(r, a_0, \theta; w_1, \omega_1\right) = V_0\left(r, a_0, \theta; w_0, \omega_0\right)$ for all r and θ.[18]

Let $V\left(r, a, \theta\right) = V_i\left(r, a, \theta; w_i, \omega_i\right)$ for each $a \in A_i$. Because V_0 and V_1 agree for action a_0, V is well defined.

In case X is trivially aligned with u on some A_i, we add $\gamma_i(a_0)^2 u(a; \theta)$ to the expression for $V_i(r, a, \theta; w_i, \omega_i)$ in (11) and adjust the definition of $w_1(\theta)$ accordingly to include any such additional $u(a_0; \theta)$ terms.

To verify that V incentivizes X, it suffices to show that at any belief p at which no action in A_i is u-optimal, $\arg\max_a \max_r \mathbb{E}_p[V(r, a, \cdot)] \subseteq A_j$, where $j \neq i$. Without loss of generality, take $i = 1$ and $j = 0$, and let p_0 denote such a belief.

Note first that, by (10), the expected value from choosing an action $a \in A_i$ followed by an optimal choice of r is equal to

$$\max_r \mathbb{E}_p[V_i(r, a, \cdot)] = \gamma_i(a_0)^2 \frac{1}{2}\left(\mathbb{E}_p[u\left(a; \cdot\right)] + \mathbb{E}_p[d\left(\cdot\right)]\right)^2$$

$$+ \mathbb{E}_p[w_i(\cdot)] - \omega_i - L\gamma_i(a_0)^2 \mathbb{E}_p\left[u\left(a; \cdot\right) + d\left(\cdot\right)\right]. \quad (12)$$

It follows that the expected value from an action $b \in A_i$ is at least as large as that from an action $a \in A_i$ if and only if

$$\frac{1}{2}\left(\mathbb{E}_p[u(b; \cdot)] + \mathbb{E}_p[d(\cdot)]\right)^2 - L\mathbb{E}_p[u(b; \cdot)] \geq \frac{1}{2}\left(\mathbb{E}_p[u(a; \cdot)] + \mathbb{E}_p[d(\cdot)]\right)^2 - L\mathbb{E}_p[u(a; \cdot)],$$

---

[18]In the general case with splitting collection $\{A_1, \ldots, A_k\}$, one can recursively define each w_{i+1} and ω_{i+1} given w_i and ω_i in an analogous fashion.

which holds if and only if $\mathbb{E}_p[u(b;\cdot)] \geq \mathbb{E}_p[u(a;\cdot)]$ since the function $(x+y)^2/2 - Lx$ is increasing in x for $x + y > L$ and L satisfies $\mathbb{E}_p[u(a';\cdot)] + \mathbb{E}_p[d(\cdot)] > L$ for all actions a'.

Suppose for contradiction that there is some $b \in A_1 \setminus \{a_0\}$ such that $b \in \arg\max_a \max_r \mathbb{E}_{p_0} V(r, a, \cdot)$. It follows that

$$\max_r \mathbb{E}_{p_0} V_1(r, b, \cdot) = \max_r \mathbb{E}_{p_0} V(r, b, \cdot) \geq \max_r \mathbb{E}_{p_0} V(r, a_0, \cdot) = \max_r \mathbb{E}_{p_0} V_1(r, a_0, \cdot).$$

By the observation in the preceding paragraph, $\mathbb{E}_{p_0} u(b; \cdot) \geq \mathbb{E}_{p_0} u(a_0; \cdot)$.

Let p_1 be a belief at which b is strictly u-optimal, i.e., $\{b\} = \arg\max_a \mathbb{E}_{p_1}[u(a;\cdot)]$. By Lemma 6, action a_0 must be u-optimal at some convex combination $p = \alpha p_1 + (1-\alpha) p_0$. At the same time, the above inequalities imply that

$$\mathbb{E}_p[u(a_0;\cdot)] = \alpha \mathbb{E}_{p_1}[u(a_0;\cdot)] + (1-\alpha)\mathbb{E}_{p_0}[u(a_0;\cdot)]$$

$$< \alpha \mathbb{E}_{p_1}[u(b;\cdot)] + (1-\alpha)\mathbb{E}_{p_0}[u(b;\cdot)] = \mathbb{E}_p[u(b;\cdot)],$$

contradicting the u-optimality of a_0 at p. $\qquad\square$

APPENDIX D. PROOFS FOR SECTION 7 AND EXTENSION OF THEOREM 2

Proof of Lemma 2. Without loss of generality, let $a_0 \in C$ be such that either $\bar{\mathsf{X}}(a_0) = 0$ or $\bar{\mathsf{X}}(a_0) \notin V_C$. (If no such a_0 exists, the lemma holds trivially.)

By Lemma 1, for each $i = 1, \ldots, n$, we have $\bar{\mathsf{X}}(a_i) = \rho_i \Delta_{a_{i-1}}^{a_i} + \sigma_i \bar{\mathsf{X}}(a_{i-1})$ for some $\sigma_i \neq 0$ and some ρ_i. Iterating these equations gives

$$\bar{\mathsf{X}}(a_0) = \bar{\mathsf{X}}(a_n) = \sum_{i=1}^{n} \Gamma_i \rho_i \Delta_{a_{i-1}}^{a_i} + \Gamma_0 \bar{\mathsf{X}}(a_0),$$

where $\Gamma_i = \sigma_n \cdots \sigma_{i+1}$ and $\Gamma_n = 1$. Because $\Delta_{a_0}^{a_1} + \cdots + \Delta_{a_{n-1}}^{a_n} = 0$, we get

$$\sum_{i=1}^{n-1} (\Gamma_i \rho_i - \rho_n) \Delta_{a_{i-1}}^{a_i} + (\Gamma_0 - 1) \bar{\mathsf{X}}(a_0) = 0.$$

Since either $\bar{\mathsf{X}}(a_0) = 0$ or $\bar{\mathsf{X}}(a_0) \notin V_C$, it follows that $\sum_{i=1}^{n-1} (\Gamma_i \rho_i - \rho_n) \Delta_{a_{i-1}}^{a_i} = 0$. Internal independence implies that $\Gamma_i \rho_i = \rho_n$ for each $i = 1, \ldots, n-1$.

If $\rho_n = 0$, then $\rho_i = 0$ for each i since $\Gamma_i \neq 0$. In this case, all $\bar{\mathsf{X}}(a_i)$ are collinear, which implies that X is trivially aligned with u on C.

For the case of $\rho_n \neq 0$, first note that, by the same iteration as above, for each k,

$$\bar{\mathsf{X}}(a_k) = \sum_{i=1}^{k-1} \rho_i \frac{\Gamma_i}{\Gamma_k} \Delta_{a_{i-1}}^{a_i} + \frac{\Gamma_0}{\Gamma_k} \bar{\mathsf{X}}(a_0) = \frac{\rho_n}{\Gamma_k} \sum_{i=1}^{k-1} \Delta_{a_{i-1}}^{a_i} + \frac{\Gamma_0}{\Gamma_k} \bar{\mathsf{X}}(a_0)$$

$$= \frac{\rho_n}{\Gamma_k} (\bar{\mathsf{u}}(a_k) - \bar{\mathsf{u}}(a_0)) + \frac{\Gamma_0}{\Gamma_k} \bar{\mathsf{X}}(a_0).$$

Letting $\gamma(a_k) = \rho_k = \rho_n/\Gamma_k \neq 0$ and $\mathsf{d} = -\bar{\mathsf{u}}(a_0) + \rho_n^{-1}\Gamma_0\bar{\mathsf{X}}(a_0)$, we have $\bar{\mathsf{X}}(a_k) = \gamma(a_k)(\bar{\mathsf{u}}(a_k) + \mathsf{d}))$ for all k, and thus X is (nontrivially) aligned with u on C. $\square$

Proof of Lemma 3. Alignment with u on B implies that there exist $\gamma_0^B, \gamma_1^B \in \mathbb{R}$ and $\mathsf{d}^B \in \mathbb{R}^\Theta$ such that

$$\gamma_0^B\bar{\mathsf{X}}(a_0) - \bar{\mathsf{u}}(a_0) = \gamma_1^B\bar{\mathsf{X}}(a_1) - \bar{\mathsf{u}}(a_1) = \mathsf{d}^B.$$

Together with the analogous equation for D, we obtain

$$\gamma_0^B\bar{\mathsf{X}}(a_0) - \gamma_1^B\bar{\mathsf{X}}(a_1) = \bar{\mathsf{u}}(a_0) - \bar{\mathsf{u}}(a_1) \text{ and } \gamma_0^D\bar{\mathsf{X}}(a_0) - \gamma_1^D\bar{\mathsf{X}}(a_1) = \bar{\mathsf{u}}(a_0) - \bar{\mathsf{u}}(a_1).$$

Subtracting one equation from the other leads to

$$\left(\gamma_0^B - \gamma_0^D\right)\bar{\mathsf{X}}(a_0) - \left(\gamma_1^B - \gamma_1^D\right)\bar{\mathsf{X}}(a_1) = 0.$$

Since $\bar{\mathsf{X}}(a_0)$ and $\bar{\mathsf{X}}(a_1)$ are linearly independent, it must be that $\gamma_k^B = \gamma_k^D$ for each $k = 0, 1$. This in turn implies that $\mathsf{d}^B = \mathsf{d}^D = \mathsf{d}$.

All that remains is to show that for any other action $a \in B \cup D$, the corresponding parameters γ_a^B and γ_a^D are equal. Since $\bar{\mathsf{X}}(a)$ cannot be collinear with both $\bar{\mathsf{X}}(a_0)$ and $\bar{\mathsf{X}}(a_1)$, we can repeat the argument replacing one of a_0 or a_1 with a to obtain $\gamma_a^B = \gamma_a^D$. $\square$
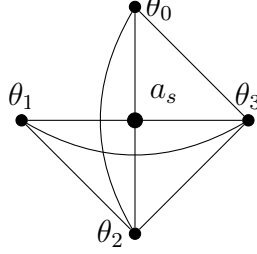
D.1. **General necessary conditions.** A set of actions $B \subseteq A$ is *cycle-rich* if it contains at least four elements and, for any proper subset $B' \subset B$ with at least three elements, there exists $a \in B \setminus B'$ such that

$$\bigcap \{V_C : C \text{ is internally independent}, a \in C, \text{ and } |C \cap B'| \geq 2\} = \{0\}.$$

The intersection above goes over all internally independent cycles that contain a and at least two elements of B'.

Example 5. Consider a variant of Example 3 in which there is an additional safe action. Thus $A = \Theta \cup \{a_s\}$ with $u(a; \theta) = r_\theta \cdot \mathbb{1}\{a = \theta\} + s \cdot \mathbb{1}\{a = a_s\}$. Suppose in addition that $\Theta = \{\theta_0, \theta_1, \theta_2, \theta_3\}$, with $r_\theta = 1/2$ for $\theta = \theta_0, \theta_1$ and $r_\theta = 1$ for

$\theta = \theta_2, \theta_3$. Let $s = 3/10$. Then the adjacency graph—which is depicted below—is incomplete since actions θ_0 and θ_1 are not adjacent, but A is cycle-rich.

Theorem 4. *Suppose $B \subseteq A$ is cycle-rich. If X is incentivizable, then it is aligned with u on B.*

Proof. We begin with the following observation.

Lemma 7. *If X is nontrivially aligned with u on B, then for any $a, b \in B$, if $\bar{\mathsf{X}}(a)$ and $\bar{\mathsf{X}}(b)$ are collinear, they are also collinear with Δ_a^b.*

Proof. Let $a, b \in B$ be such that $\bar{\mathsf{X}}(a)$ and $\bar{\mathsf{X}}(b)$ are collinear and let $\gamma(\cdot) \neq 0$ and d be such that $\bar{\mathsf{X}}(a) = \gamma(a)\left(\bar{\mathsf{u}}(a) + \mathsf{d}\right)$ and $\bar{\mathsf{X}}(b) = \gamma(b)\left(\bar{\mathsf{u}}(b) + \mathsf{d}\right)$. By the collinearity assumption, there exists $\alpha \neq 0$ such that $\bar{\mathsf{u}}(a) + \mathsf{d} = \alpha\left(\bar{\mathsf{u}}(b) + \mathsf{d}\right)$. Because $\bar{\mathsf{u}}(a) \neq \bar{\mathsf{u}}(b)$, it must be that $\alpha \neq 1$. It follows that $\mathsf{d} = \frac{1}{1-\alpha}\left(\alpha\bar{\mathsf{u}}(b) - \bar{\mathsf{u}}(a)\right)$ and

$$\frac{1}{\gamma(a)}\bar{\mathsf{X}}(a) = \bar{\mathsf{u}}(a) + \frac{1}{1-\alpha}\left(\alpha\bar{\mathsf{u}}(b) - \bar{\mathsf{u}}(a)\right) = \frac{\alpha}{1-\alpha}\left(\bar{\mathsf{u}}(b) - \bar{\mathsf{u}}(a)\right),$$

as needed. $\square$

Cycle-richness implies that there is some $a \in B$ and a collection of internally independent cycles $\tilde{C}$ with $a \in \tilde{C} \subseteq B$ for which the intersection of the spaces $V_{\tilde{C}}$ is $\{0\}$. Thus either $\bar{\mathsf{X}}(a) = 0$ or $\bar{\mathsf{X}}(a) \notin V_C$ for some such cycle C. By Lemma 2, X is aligned with u on C.

Let $B' \subseteq B$ be a subset of B of maximal cardinality on which X is aligned with u. By the above argument, B' has at least three elements. Suppose for contradiction that $B' \neq B$. By the same argument as in the preceding paragraph, cycle-richness implies that there exists $a \in B \setminus B'$ and a cycle C containing a such that $|C \cap B'| \geq 2$ and either $\bar{\mathsf{X}}(a) = 0$ or $\bar{\mathsf{X}}(a) \notin V_C$. By Lemma 2, X is aligned with u on C.

If there exists a pair of distinct actions $a_0, a_1 \in C \cap B'$ such that $\bar{\mathsf{X}}(a_0)$ and $\bar{\mathsf{X}}(a_1)$ are not collinear, Lemma 3 implies that X is aligned with u on $C \cup B'$, contradicting the maximality of B'.

From now on, suppose that $a_0, a_1 \in C \cap B'$ are distinct actions such that $\bar{\mathsf{X}}(a_0)$ and $\bar{\mathsf{X}}(a_1)$ are collinear.

If $\bar{\mathsf{X}}(a_0)$ and $\bar{\mathsf{X}}(a_1)$ are not collinear with $\Delta_{a_0}^{a_1}$, then Lemma 7 implies that the alignment with u on C and that on B' must both be trivial, which further implies that, for all $b \in C \cup B'$, $\bar{\mathsf{X}}(b)$ is collinear with $\bar{\mathsf{X}}(a_0)$. Thus, X is trivially aligned with u on $C \cup B'$, contradicting the maximality of B'.

If $\bar{\mathsf{X}}(a_0)$ and $\bar{\mathsf{X}}(a_1)$ are collinear with $\Delta_{a_0}^{a_1}$, then, by Lemma 1, $\bar{\mathsf{X}}(a) \in \mathrm{span}(\Delta_a^{a_0}, \Delta_{a_0}^{a_1}) = V_C$. The choice of cycle C implies that $\bar{\mathsf{X}}(a) = 0$. Another application of Lemma 1 shows that $\bar{\mathsf{X}}(a_0)$ is collinear with $\Delta_a^{a_0}$, which contradicts collinearity with $\Delta_{a_0}^{a_1}$ due to the independence assumptions. $\qquad\square$

Proof of Theorem 2. It suffices to show that the set of all actions is cycle-rich; the result then follows from Theorem 4.

Take any proper subset $B \subset A$ with at least three actions $b_0, b_1, b_2 \in B$ and let $a \in A \setminus B$. Consider cycles $C_i = B \setminus \{b_i\} \cup \{a\}$. Then, $V_{C_i} = \mathrm{span}\{\Delta_a^{b_j} : j \neq i\}$. The independence assumption implies that $\bigcap_i V_{C_i} = \{0\}$. $\qquad\square$

It is straightforward to extend Theorem 4 to problems in which there is a splitting collection $\{A_0, \ldots, A_k\}$ such that, for each l, either A_l is cycle-rich or it contains exactly two elements. In that case, only questions piecewise aligned with u are incentivizable.

APPENDIX E. PROOF OF THEOREM 3

This appendix is divided into the following subsections. Section E.1 shows that each incentivizable question in a product problem can be decomposed into linearly independent vectors that correspond to different tasks. We use this decomposition together with Lemma 1 to derive restrictions on questions for adjacent actions. Subsections E.2 to E.5 show that all adjacency cycles are exact. Section E.3 develops useful tools, and Sections E.4 and E.5 deal with different classes of cycles. Section E.6 concludes the proof.

E.1. **Decomposition.** Assume throughout that X is incentivizable.

For each i and each $t_i \in \Theta_i$, let $\mathsf{e}_i(t_i) \in \mathbb{R}^\Theta$ be the vector such that, for each $\theta \in \Theta$, $\mathsf{e}_i(\theta|t_i) = \mathbf{1}\{\theta_i = t_i\}$. Let $E_i = \mathrm{span}\left\{\Delta_{b_i}^{a_i} : a_i, b_i \in A_i\right\} \subseteq \mathbb{R}^\Theta$ be the subspace of $\mathbb{R}^\Theta$ spanned bu the utility difference vectors (see also footnote 15). Because payoffs in task i depend only on the state coordinate i, subspaces E_i are linearly independent. Let E_0 be a complementary space to the sum of E_1 through E_I.

For each a, $\bar{\mathsf{X}}(a) = \sum_{i=0}^{I} \mathsf{w}_i(a)$ admits a unique decomposition to $\mathsf{w}_i(a) \in E_i$ for all i. Note that the vectors $\mathsf{w}_0(a), \mathsf{w}_1(a), \ldots, \mathsf{w}_I(a)$ are linearly independent.

Take any distinct a and b such that $a_{-i} = b_{-i}$ for some i. Then a and b are adjacent, and, by Lemma 1, there are $x(a,b) \neq 0$ and $y(a,b)$ such that $\bar{\mathsf{X}}(a) = x(a,b)\bar{\mathsf{X}}(b) + y(a,b)\Delta_b^a$. Using the above decomposition, we have

$$0 = \left[\mathsf{w}_i(a) - x(a,b)\mathsf{w}_i(b) - y(a,b)\Delta_{b_i}^{a_i} \right] + \sum_{j \neq i} \left[\mathsf{w}_j(a) - x(a,b)\mathsf{w}_j(b) \right] + \left[\mathsf{w}_0(a) - x(a,b)\mathsf{w}_0(b) \right].$$

The proof of Lemma 1 shows that $x(a,b)$ is uniquely defined if and only if $\bar{\mathsf{X}}(a) \neq 0$ and $\bar{\mathsf{X}}(a)$ is not collinear with $\Delta_{b_i}^{a_i}$. When this is not the case, we call (a,b) a *free transition*. The values of $x(a,b)$ for free transitions are carefully chosen below. Our choice always satisfies $x(a,b)x(b,a) = 1$ (which always holds for non-free transitions). Let $x(a,a) = 1$.

Because all vectors in square brackets lie in linearly independent spaces, they must all be equal to 0, and hence

$$\mathsf{w}_i(a) - x(a,b)\mathsf{w}_i(b) = y(a,b)\Delta_{b_i}^{a_i}, \tag{13}$$

$$\mathsf{w}_j(a) - x(a,b)\mathsf{w}_j(b) = 0 \text{ for each } j \neq i, \tag{14}$$

$$\text{and} \quad \mathsf{w}_0(a) - x(a,b)\mathsf{w}_0(b) = 0. \tag{15}$$

Lemma 8. *There exist a vector w_0^*, vectors $\mathsf{w}_i^*(a_i)$ for each $i = 1, \ldots, I$ and a_i, and scalars $\gamma_i(a) \neq 0$ for each a and $i = 0, \ldots, I$ such that $\mathsf{w}_i(a) = \gamma_i(a)\mathsf{w}_i^*(a_i)$ for each i, and $\mathsf{w}_0(a) = \gamma_0(a)\mathsf{w}_0^*$.*

This result says that, for a fixed a_i, the vectors $\mathsf{w}_i(a_{-i}a_i)$ for a_{-i} are either all collinear or all equal to 0.

Proof. For the first claim, fix an action a^*. For each a_i, let $\mathsf{w}_i^*(a_i) = \mathsf{w}_i(a_{-i}^* a_i)$. For each a, fix an arbitrary a path of adjacent actions $a^0 = a_{-i}^* a_i, \ldots, a^n = a$ such that for each $l < n$, $a_i^l = a_i$. Let $\gamma_i(a) = x(a^n, a^{n-1}) \cdots x(a^1, a^0)$. A repeated application of (14) shows that $\mathsf{w}_i(a) = \gamma_i(a)\mathsf{w}_i^*(a_i)$. Analogous argument shows the second claim. $\square$

Lemma 9. *For each i, one of the following is true:*
  (i) $\mathsf{w}_i^*(a_i) = 0$ for each a_i,
  (ii) there exists $a_i^0 \in A_i$ such that $\mathsf{w}_i^*(a_i^0) = 0$ and $\mathsf{w}_i^*(a_i) \neq 0$ for each $a_i \neq a_i^0$,
  (iii) $\mathsf{w}_i^*(a_i) \neq 0$ for each a_i.
If question X does not depend trivially on task i, then there is an action a_i such that $\mathsf{w}_i^*(a_i) \neq 0$.

Proof. Suppose that there are three different actions a_i, b_i, and c_i such that $\mathsf{w}_i^*(a_i) = \mathsf{w}_i^*(b_i) = 0$ and $\mathsf{w}_i^*(c_i) \neq 0$. Equation (13) together with Lemma 8 imply that, for any c_{-i}, $\mathsf{w}_i(c_i c_{-i})$ is simultaneously collinear with $\Delta_{c_i}^{a_i}$ and $\Delta_{c_i}^{b_i}$, contradicting the independence assumption.

For the last claim, suppose that $\mathsf{w}_i^*(a_i) = 0$ for each a_i. Equation (13) together with Lemma 8 imply that $y(a, b) = 0$ for each a and b such that $a_{-i} = b_{-i}$. But then, for each a_{-i}, the vectors $\bar{\mathsf{X}}(a_{-i} a_i)$ and $\bar{\mathsf{X}}(a_{-i} b_i) = x(a_{-i} b_i, a_{-i} a_i) \bar{\mathsf{X}}(a_{-i} a_i)$ are collinear. Hence, X depends on task i trivially. $\qquad\square$

Case (i) of Lemma 9 is equivalent to X depending on problem i trivially. Because X depends nontrivially on at least three tasks, there must be at least three tasks in cases (ii) or (iii).

E.2. **Exact cycles.** Recall that an adjacency cycle $a^0, \ldots, a^n = a^0$ is *exact* if $\prod_{l<n} x(a^l, a^{l+1}) = 1$. The goal of this subsection as well as subsections E.3 to E.5 is to prove the following result.

Lemma 10. *The values $x(a, b)$ for free transitions (a, b) can be chosen so that (i) $x(a, b) x(b, a) = 1$ for all adjacent a and b, and (ii) every adjacency cycle is exact.*

Lemma 8 and a repeated application of equation (15) shows that, if $\mathsf{w}_0^* \neq 0$, then, all adjacency cycles are exact. From now on, we assume that $\mathsf{w}_0^* = 0$.

E.3. **Tools.** The two results in this section develop tools that we use in the subsequent analysis.

The first tool allows us to replace the problem of whether a cycle is exact with a problem about related cycles. For each path $c = (a^0, \ldots, a^{n_c})$ and each transition (a, b) between two adjacent actions, define

$$m_a^b(c) = \# \left\{ l < n_c : a^l = a, a^{l+1} = b \right\} - \# \left\{ l < n_c : a^l = b, a^{l+1} = a \right\}.$$

Lemma 11. *Suppose that, for some adjacency cycle c, there exists a collection D of exact adjacency cycles such that for each adjacent pair a, b, $m_a^b(c) = \sum_{d \in D} m_a^b(d)$. Then cycle c is exact.*

Proof. Let $\prec$ be an arbitrary strict order on the set of actions A. Then we have

$$\prod_{l<n_c} x(a^{c,l}, a^{c,l+1}) = \prod_{a \prec b} (x(a, b))^{m_a^b(c)} = \prod_{a \prec b} (x(a, b))^{\sum_{d \in D} m_a^b(d)}$$

$$= \prod_{d \in D} \prod_{a \prec b} (x(a, b))^{m_a^b(d)} = \prod_{d \in D} \prod_{l<n_d} x(a^{d,l}, a^{d,l+1}) = 1,$$

where the last equality comes from the fact that cycles in D are exact. $\qquad\square$

The second result shows that it is enough to consider particular kinds of "small" cycles.

Lemma 12. *Every adjacency cycle is exact if (and only if) the following cycles are exact:*

(1) (a, b, a) *for all adjacent actions a and b,*

(2) (a, b, c, a) *for all triples of actions a, b, and c such that $a_{-i} = b_{-i} = c_{-i}$,*

(3) $(a, a_{-i}b_i, a_{-ij}b_ib_j, a_{-j}b_j, a)$ *for all actions a, $i \neq j$, and b_i and b_j.*

Proof. Call cycles of the three forms described in the lemma *small* and any other cycle *large*.

Take any large cycle of adjacent actions $a = a^0, \ldots, a^n = a$ and let i_l be such that, for each $l < n$, $a^l_{-i_l} = a^{l+1}_{-i_l}$. For future reference, notice that if the action in some task i is ever changed, then it must be changed at least twice: if $i_l = i$ for some l, then there is some $l' \neq l$ such that $i_{l'} = i$. We use Lemma 11 and small cycles to reorder and reduce the large cycle without changing the value of the product of the associated x terms:

- if $i_l > i_{l+1}$ for $l < n - 1$ we use the small cycle of type (3) to switch the order of the two tasks, i.e., replace the cycle fragment $\ldots, a^l, a^{l+1}, a^{l+2}, \ldots$, where $a^{l+1} = a^l_{-i_l}a^{l+2}_{i_l}$, with $\ldots, a^l, a^l_{-i_{l+1}}a^{l+2}_{i_{l+1}}, a^{l+2}, \ldots$. More precisely, due to Lemma 11 and the fact that the small cycle is exact, the exactness of the original cycle is equivalent to the exactness of the modified cycle;

- if $i_l = i_{l+1}$ (including possibly $i_{n-1} = i_0$), we use either type (1) or type (2) to reduce the large cycle, i.e, replace the cycle fragment $\ldots, a^l, a^{l+1}, a^{l+2}, \ldots$ with $\ldots, a^l, a^{l+2}, \ldots$ in the case of a type (2) cycle or with $\ldots, a^l, \ldots$ in the case of a type (1) cycle.

Consider a process in which one of the above operations is applied until it cannot be applied anymore. Because either the operations reorder tasks in an increasing direction, or, if they don't change the order, they reduce the length of the cycle, the process never reverts and it will eventually stop. If the process stops at a single-element cycle a, then, because $x(a, a) = 1$, the original cycle must be exact.

Otherwise, the process stops with a nontrivial cycle $a = a^0, \ldots, a^m = a$ for some $2 \leq m \leq n$. Then it must be that $i_l < i_l + 1$ for each $l < m$. But this contradicts the earlier observation that, in an adjacency cycle, if i appears at least once, it must appear at least twice. $\qquad\square$

E.4. **Cycles without free transitions.** Here, we consider the exactness of small cycles without free transitions. We refer to i from the definition of type (2) cycles as the *relevant task* for this cycle; similarly, we refer to i and j as the relevant tasks for type (3) cycles. We say that a type (2) or type (3) cycle $(a^0, \dots, a^n = a^0)$ is *grounded* if there exists some k such that k is not a relevant task and $\mathsf{w}_k^*(a_k^0) \neq 0$.

Lemma 13. *Any grounded type (2) or type (3) cycle is exact.*

Proof. Suppose that $\mathsf{w}_k^*(a_k^0) \neq 0$ for some irrelevant task k. Then $\mathsf{w}_k(a^i) \neq 0$ for each action a^i in the cycle. A repeated application of (14) shows that for each $l \leq n$,

$$\mathsf{w}_k(a^l) = \mathsf{w}_k(a^0)x(a^0, a^1) \cdots x(a^{l-1}, a^l).$$

The result follows from the fact that $a^n = a^0$. $\qquad\square$

Lemma 14. *Let $(a, a_{-i}b_i, a_{-ij}b_ib_j, a_{-j}b_j, a)$ be a type (3) small cycle such that either (i) $\mathsf{w}_i^*(a_i) \neq 0$ and $\mathsf{w}_i^*(a_i)$ is not collinear with $\Delta_{a_i}^{b_i}$, or (ii) $\mathsf{w}_j^*(a_j) \neq 0$ and $\mathsf{w}_j^*(a_j)$ is not collinear with $\Delta_{a_j}^{b_j}$. Then the cycle is exact.*

Proof. Using 13 and 14, we get

$$x(a_{-ij}b_ib_j, a_{-i}b_i)\left(x(a_{-i}b_i, a)\mathsf{w}_i(a) + y(a_{-i}b_i, a)\Delta_{a_i}^{b_i}\right)$$
$$= x(a_{-ij}b_ib_j, a_{-i}b_i)\mathsf{w}_i(a_{-i}b_i) = \mathsf{w}_i(a_{-ij}b_ib_j)$$
$$= x(a_{-ij}b_ib_j, a_{-j}b_j)\mathsf{w}_i(a_{-j}b_j) + y(a_{-ij}b_ib_j, a_{-j}b_j)\Delta_{a_i}^{b_i}$$
$$= x(a_{-ij}b_ib_j, a_{-j}b_j)x(a_{-j}b_j, a)\mathsf{w}_i(a) + y(a_{-ij}b_ib_j, a_{-j}b_j)\Delta_{a_i}^{b_i}.$$

Suppose without loss of generality that $\mathsf{w}_i^*(a_i) \neq 0$ and $\mathsf{w}_i^*(a_i)$ is not collinear with $\Delta_{a_i}^{b_i}$, which implies that $\mathsf{w}_i(a) \neq 0$ and $\mathsf{w}_i(a)$ is not collinear with $\Delta_{a_i}^{b_i}$. The first and the last line of the above sequence of equalities yield

$$x(a_{-ij}b_ib_j, a_{-i}b_i)x(a_{-i}b_i, a) = x(a_{-ij}b_ib_j, a_{-j}b_j)x(a_{-j}b_j, a).$$

Hence, the cycle is exact. $\qquad\square$

Lemma 15. *Suppose that a type (2) or type (3) small cycle is such that $\mathsf{w}_i^*(a_i) \neq 0$ for each action a in the cycle and each relevant task i and none of the transitions is free. Then it is exact.*

Proof. If the cycle is grounded, the result follows from Lemma 13. Accordingly, suppose the cycle is not grounded. Then, because of Lemma 9, and due to the
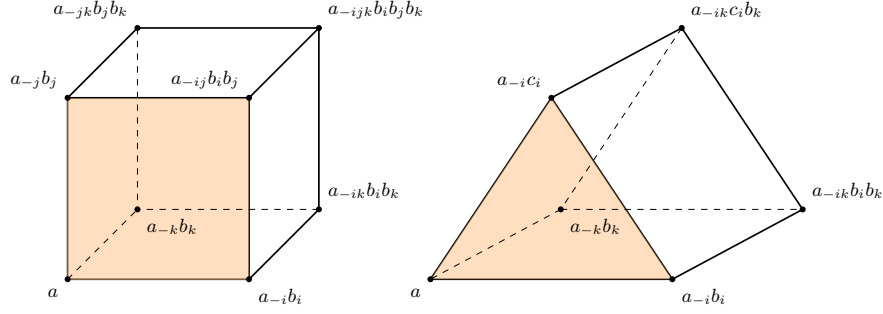
FIGURE 4. Small cycles of type (3) and type (2)

assumption that at least three different tasks are nontrivial, there exists a non-relevant k and an action $b_k \neq a_k$ such that $\mathsf{w}_k^*(b_k) \neq 0$.

Suppose that the original cycle $(a, a_{-i}b_i, a_{-ij}b_ib_j, a_{-j}b_j, a)$ is of type (3). Such a cycle corresponds to the orange face on the left-side of Figure 4. Consider the type (3) cycles that are associated with all five remaining faces of the cube:

- the bottom, $(a, a_{-i}b_i, a_{-ik}b_ib_k, a_{-k}b_k, a)$;
- the right, $(a_{-i}b_i, a_{-ij}b_ib_j, a_{-ijk}b_ib_jb_k, a_{-ik}b_ib_k, a_{-i}b_i)$;
- the top, $(a_{-j}b_j, a_{-jk}b_jb_k, a_{-ijk}b_ib_jb_k, a_{-ij}b_ib_j, a_{-j}b_j)$;
- the left, $(a, a_{-k}b_k, a_{-jk}b_jb_k, a_{-j}b_j, a)$;
- and the back, $(a_{-k}b_k, a_{-ik}b_ib_k, a_{-ijk}b_ib_jb_k, a_{-jk}b_jb_k, a_{-k}b_k)$.

All five of these cycles are grounded (by the assumption that $\mathsf{w}_i^*(a_i) \neq 0$ for each action a in the cycle and each relevant task i), and hence exact by Lemma 13. Moreover, the conditions of Lemma 11 are satisfied. Therefore, the original cycle is exact as well.

If the cycle is of type (2), then because the cycle is not grounded, we have $\bar{\mathsf{X}}(a) = \gamma_i(a)\mathsf{w}_i^*(a_i)$, $\bar{\mathsf{X}}(a_{-i}b_i) = \gamma_i(a_{-i}b_i)\mathsf{w}_i^*(b_i)$, and $\bar{\mathsf{X}}(a_{-i}c_i) = \gamma_i(a_{-i}c_i)\mathsf{w}_i^*(c_i)$. Because the transition $(a, a_{-i}b_i)$ is not free, the vector $\mathsf{w}_i^*(a_i)$ cannot be collinear with $\Delta_{a_i}^{b_i}$ (otherwise, $\bar{\mathsf{X}}(a)$ would also be collinear with $\Delta_{a_i}^{b_i}$, in which case the transition is free). As a result, Lemma 14 applies to the 4-action cycle $(a, a_{-i}b_i, a_{-ik}b_ib_k, a_{-k}b_k, a)$, which means that this cycle is exact; see the right panel of Figure 4. An analogous observation holds for the other two 4-action cycles depicted in Figure 4. The 3-action cycle in the back is grounded, and hence exact by Lemma 13. Finally, an application of Lemma 11 shows that the original cycle is exact as well. $\qquad\square$

Lemma 16. *If a type (2) or (3) cycle has no free transitions, then it is exact.*

Proof. By Lemmas 13 and 15, it suffices to assume that the cycle is not grounded and $w_i^*(a_i) = 0$ for some relevant i and action a in the cycle. In such a case, if the cycle were of type (2), all transitions to action a would be free.

Suppose the cycle is of type (3). Let $b \neq a$ be the action in the cycle such that $a_{-j} = b_{-j}$. It follows that $w_k^*(a_i) \neq 0$ if and only if $k = j$. Then

$$\bar{X}(a) = \gamma_j(a)w_j^*(a_j) \text{ and } \bar{X}(b) = \gamma_j(b)w_j^*(b_j) = x(b,a)\gamma_j(a)w_j^*(a_j) + y(b,a)\Delta_{a_j}^{b_j}.$$

Because the transition (a, b) is not free, it must be that $w_j^*(a_j) \neq 0$ and $w_j^*(a_j)$ is not collinear with $\Delta_{a_j}^{b_j}$. Lemma 14 therefore implies that the cycle is exact. $\qquad \square$

E.5. **Cycles with free transitions.** In this subsection, we show how to determine values of $x(a, b)$ for free transitions so that all cycles that include such transitions are exact. For simplicity, if X depends trivially on task i, we say simply that i is trivial (and similarly for nontrivial dependence).

Notice first that, if a transition between adjacent actions a and b such that $a_{-i} = b_{-i}$ is free, then it must be that $w_j^*(a_j) = 0$ for each $j \neq i$. Indeed, if, for some $\alpha \neq 0$,

$$\bar{X}(a) = \alpha\bar{X}(b) = x(a,b)\bar{X}(b) + y(a,b)\Delta_{a_i}^{b_i},$$

then $\bar{X}(b)$, and hence $\bar{X}(a)$, must be collinear with $\Delta_{a_i}^{b_i}$. Because of the linear independence of $\Delta_{a_i}^{b_i}$ and $w_j(a)$ for each $j \neq i$, it must be that $w_j(a) = 0$, and hence $w_j^*(a_j) = 0$. We refer to this property as the *test* for freeness of the transition (which provides necessary conditions). It follows that, if $j \neq i$ is nontrivial, then task j is in case (ii) from Lemma 9 and $a_j = a_j^0$.

In what follows, we consider two cases: *Case I*: There exists a single task in case (iii) from Lemma 9, i.e., one nontrivial i such that $w_i^*(a_i) \neq 0$ for all a_i. Assume without loss of generality that $i = I$. In this case, all free transitions must be between adjacent actions a and b such that $a_{-I} = b_{-I}$. Moreover, it must be that $w_I^*(a_I)$ and $w_I^*(b_I)$ are collinear with $\Delta_{a_I}^{b_I}$.

Assume without loss of generality that task 1 is nontrivial and fix action $b_1 \neq a_1$ so that $w^*(b_1) \neq 0$. Let

$$x(a, b) = x(a, a_{-1}b_1)x(a_{-1}b_1, a_{-1I}b_1b_I)x(a_{-1I}b_1b_I, b).$$

The above definition implies that the cycle $(a, a_{-1}b_1, a_{-1I}b_1b_I, b, a)$ is exact. This cycle corresponds to the red face(s) in Figure E.5. The orange edge corresponds to the free transition. Notice that each red face cycle contains only one free transition. This is

because of the test: all other transitions of the cycle either keep fixed the action in task I or the action b_1, and those actions are associated with nonzero $\mathsf{w}^*_\cdot(\cdot)$ vectors.
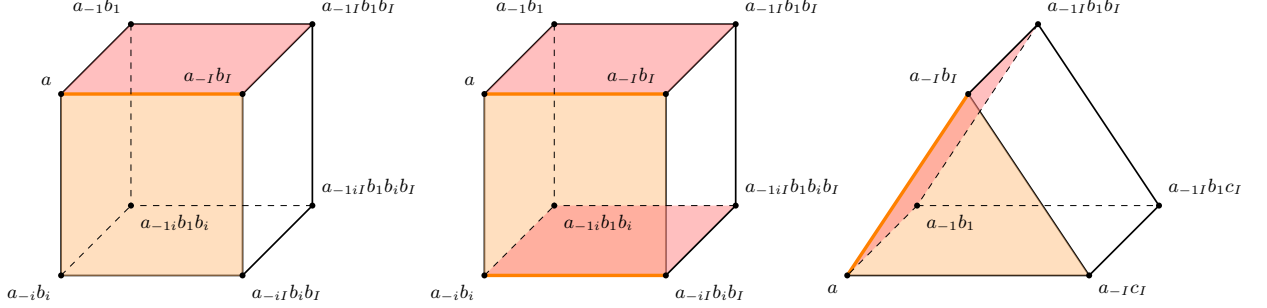

FIGURE 5. Cycles with free transitions in case I.

There are three types of cycles that contain free transitions, depicted as orange faces in Figure E.5.

The left panel corresponds to the cycle $(a, a_{-i}b_i, a_{-iI}b_ib_I, a_{-I}b_I, a)$ when task i is nontrivial. In this case, $b_i \neq a_i^0$, and, by Lemma 9, $\mathsf{w}^*_i(b_i) \neq 0$. The test implies that none of the other transitions in the orange cycle are free: either action b_i or action a_I is fixed. Analogously, an application of the test shows that none of the other cycles (the uncolored faces) is free: one of the actions a_I, b_i, or b_1 is fixed. Proceeding as in the proof of Lemma 15, we see that this cycle is exact.

The center panel corresponds to a situation when task i is trivial. In this case, the transition $(a_{-i}b_i, a_{-iI}b_ib_I)$ is free and $x(a_{-i}b_i, a_{-iI}b_ib_I)$ can be chosen to make the cycle on the bottom face exact. None of the other transitions are free. Because the red faces and the uncolored cycles are exact, the above argument implies that the orange face cycle is exact as well.

The right panel corresponds to the orange cycle $(a, a_{-I}b_I, a_{-I}c_I, a)$. The other transitions of the orange cycle are not free (otherwise, $w^*(a_I)$ would be collinear with $\Delta_{a_I}^{c_I}$, which would violate linear independence of the latter vector with $\Delta_{a_I}^{b_I}$). All other transitions fix one of the actions: a_I, b_I, c_I, or b_1. Hence, due to the test, none of the remaining transitions are free. The claim follows from the same reasoning as in Lemma 15.

Case II: For all nontrivial i, there exists a unique a_i^0 such that $\mathsf{w}^*_i(a_i^0) = 0$. Let a^0 be the product problem action that consists of actions a_i^0. Assume without loss of generality that task $i = 1$ is nontrivial. Fix action $a_1^* \neq a_1^0$.

In this case, a transition is free if and only if it takes the form $(a_{-i}^0 a_i, a^0)$ for some i. Indeed, the above observation implies that if transition (a, b) is free, then

$a_{-i} = a^0_{-i}$. Furthermore, if neither $a = a^0$ nor $b = a^0$, then both $w^*_i(a_i)$ and $w^*_i(b_i)$ must be collinear with $\Delta^{b_i}_{a_i}$. But, together with the linear independence assumption, this implies that $w^*_i(a_i)$ is not collinear with $\Delta^{a^0_i}_{a_i}$, which contradicts the fact that

$$x(a^0, a)\gamma_i(a)w^*_i(a_i) + y(a^0, a)\Delta^{a^0_i}_{a_i} = \gamma_i(a^0)w^*_i(a^0_i) = 0.$$

For each $b_1 \neq a^*_1$, each $i \neq 1$, and each b_i, let

$$x(a^0_{-1}a^*_1, a^0) = 1, \ x(a^0_{-1}b_1, a^0) = x(a^0_{-1}b_1, a^0_{-1}a^*_1), \ \text{and}$$

$$x(a^0_{-i}b_i, a^0) = x(a^0_{-i}b_i, a^0_{-1i}a^*_1 b_i)x(a^0_{-1i}a^*_1 b_i, a^0_{-1}a^*_1).$$

Due to the above definition, the cycles $(a^0, a^0_{-1}b_1, a^0_{-1}a^*_1, a^0)$ and $(a^0, a^0_{-i}b_i, a^0_{-i1}a^*_1 b_i, a^0_{-1}a^*_1, a^0)$ are exact. These cycles correspond to the red faces in Figure E.5.

There are three types of cycles that contain free transitions in this case other than the cycles listed above.

First, consider a cycle $(a^0, a^0_{-i}b_i, a^0_{-ij}b_i b_j, a^0_{-j}b_j, a^0)$ for $i \neq 1$. This cycle is depicted in orange on the left panel of Figure E.5. The two cycles depicted in red are exact due to the choice of the x coefficients. Finally, all of the other cycles are exact because they do not contain free transitions.
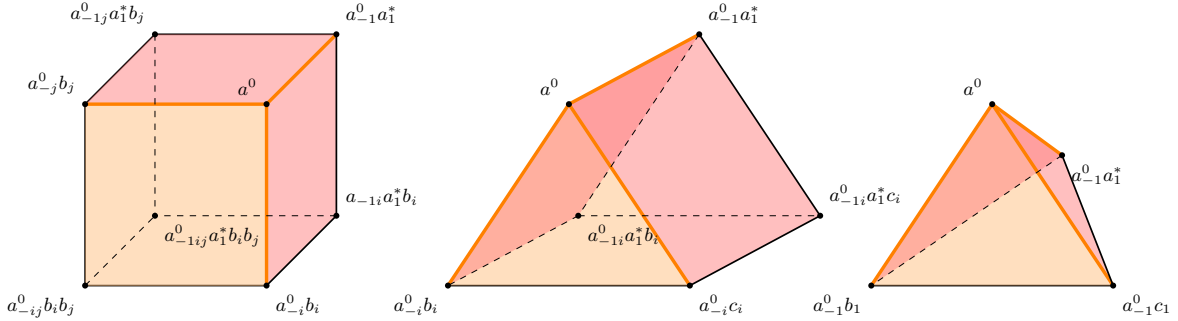

FIGURE 6. Cycles with free transitions in case II.

Second, consider a cycle $(a^0, a^0_{-i}b_i, a^0_{-i}c_i, a^0)$ for $i \neq 1$. This cycle corresponds to front wall depicted in orange on the center panel of Figure E.5. The two cycles depicted in red are exact due to the choice of the x coefficients. All of the other cycles (corresponding to the back and bottom walls) are exact because they do not contain free transitions.

Third, consider a cycle $(a^0, a^0_{-1}b_1, a^0_{-1}c_1, a^0)$. This cycle corresponds to the front wall depicted in orange on the left panel of Figure E.5. The two cycles depicted in red (corresponding to the left and right walls) are exact due to the choice of the x

coefficients. The remaining cycle (corresponding to the bottom wall) is exact because it does not contain any free transitions.

Proceeding as in the proof of Lemma 15, we see that each of the considered cycles is exact.

Because there can be at most one task in case (iii) from Lemma 9, there are no other cases to consider. Together with Lemma 16, this concludes the proof of Lemma 10.

E.6. **Conclusion of the proof of Theorem 3.** Recall that $\mathsf{w}_i(a) = \gamma_i(a)\mathsf{w}_i^*(a_i)$. The next result delivers additional information about the function $\gamma_i(\cdot)$.

Lemma 17. *There exist* $\gamma : A \longrightarrow \mathbb{R}$, $\gamma_i^* : A_i \longrightarrow \mathbb{R}$, *and* $\gamma_0^* \in \mathbb{R}$ *such that, for each* a, $\gamma_i(a) = \gamma(a)\gamma_i^*(a_i)$ *for each* $i = 1, \ldots, I$, *and* $\gamma_0(a) = \gamma(a)\gamma_0^*$. *In addition,* $x(a, b) = \frac{\gamma(a)}{\gamma(b)}$ *for any two adjacent actions* a *and* b.

Proof. Fix an action a^*. For each a, find a path of adjacent actions $a^* = a^0, \ldots, a^m = a$. Define $\gamma(a) = x(a^m, a^{m-1}) \cdots x(a^1, a^0)$. Lemma 10 implies that $\gamma(a)$ is well defined in that its definition does not depend on the choice of path from a^* to a. Moreover, for any two adjacent actions a and b, if $a^* = a^0, \ldots, a^m = a$ is an adjacency path from a^* to a, then $a^* = a^0, \ldots, a^m, b$ is an adjacency path from a^* to b, and

$$\gamma(b) = x(b, a)x(a^m, a^{m-1}) \cdots x(a^1, a^0) = x(b, a)\gamma(a).$$

Let $\gamma_i^*(a_i) = \gamma_i(a_{-i}^* a_i)/\gamma(a_{-i}^* a_i)$. The claim for $i > 0$ follows from the fact that, for each a, if $\mathsf{w}_i^*(a_i) \neq 0$, then

$$\mathsf{w}_i(a) = \frac{\gamma_i(a)}{\gamma_i(a_{-i}^* a_i)}\mathsf{w}_i(a_{-i}^* a_i) = \frac{\gamma(a)}{\gamma(a_{-i}^* a_i)}\mathsf{w}_i(a_{-i}^* a_i).$$

A similar argument establishes the claim for $i = 0$ (see also the proof of Lemma 8). $\qquad\square$

Lemma 18. *There exist* $y_i^* \in \mathbb{R}$ *and* $\mathsf{d}_i \in E_i$ *such that* $\gamma_i^*(a_i)\mathsf{w}_i^*(a_i) = y_i^*\bar{\mathsf{u}}_i(a_i) + \mathsf{d}_i$ *for any* a_i.

Proof. By (13), for any actions a and b such that $a_{-i} = b_{-i}$,

$$\gamma_i^*(a_i)\mathsf{w}_i^*(a_i) - \gamma_i^*(b_i)\mathsf{w}_i^*(b_i) = \frac{1}{\gamma(a)}\mathsf{w}_i(a) - \frac{1}{\gamma(b)}\mathsf{w}_i(b) \tag{16}$$

$$= \frac{1}{\gamma(a)}\left(\mathsf{w}_i(a) - x(a, b)\mathsf{w}_i(b)\right) = \frac{1}{\gamma(a)}y(a, b)\Delta_{b_i}^{a_i}.$$

Because the left-hand side and $\Delta_{b_i}^{a_i}$ do not depend on a_{-i}, neither does $y(a,b)/\gamma(a)$. Let $y_i^*(a_i, b_i) = y(a,b)/\gamma(a)$.

If task i has only two actions, it is easy to see that $y_i^*(a_i, b_i) = y_i^*(b_i, a_i) =: y_i^*$. The claim follows.

If task i has at least three actions, take a, b, and c such that $a_{-i} = b_{-i} = c_{-i}$ and a_i, b_i, and c_i are distinct. Applying the above equation to pairs (a, b), (b, c), and (c, a) yields

$$y_i^*(a_i, c_i)\Delta_{b_i}^{a_i} + y_i^*(a_i, c_i)\Delta_{c_i}^{b_i} = y_i^*(a_i, c_i)\Delta_{c_i}^{a_i} = \gamma_i^*(a_i)\mathsf{w}_i^*(a_i) - \gamma_i^*(c_i)\mathsf{w}_i^*(c_i)$$
$$= \gamma_i^*(a_i)\mathsf{w}_i^*(a_i) - \gamma_i^*(b_i)\mathsf{w}_i^*(b_i) + \gamma_i^*(b_i)\mathsf{w}_i^*(b_i) - \gamma_i^*(c_i)\mathsf{w}_i^*(c_i)$$
$$= y_i^*(a_i, b_i)\Delta_{b_i}^{a_i} + y_i^*(b_i, c_i)\Delta_{c_i}^{b_i}.$$

The independence assumption implies that $y_i^*(a_i, b_i) = y_i^*(a_i, c_i)$. Because the claim holds for arbitrary and distinct actions, there must be y_i^* such that for all a_i and b_i, we have $y_i^*(a_i, b_i) = y_i^*$.

Finally, fix a_i^* and take $\mathsf{d}_i = \gamma_i^*(a_i)\mathsf{w}_i^*(a_i) - y_i^*\bar{\mathsf{u}}_i(a_i^*)$. The claim follows from equation (16). $\qquad\square$

Substituting the observations from the two lemmas back into the decomposition of $\bar{\mathsf{X}} = \sum_{i=0}^n \mathsf{w}_i(a)$, we obtain

$$\bar{X}(a) = \gamma(a)\left(\sum_{i=1}^n \gamma_i^*(a_i)\mathsf{w}_i^*(a_i) + \gamma_0\mathsf{w}_0^*\right) = \gamma(a)\left(\sum_{i=1}^n y_i^*\bar{\mathsf{u}}_i(a_i) + \left[\sum_{i=1}^n \mathsf{d}_i + \gamma_0\mathsf{w}_0^*\right]\right).$$

Let d be the vector in the square brackets. The result follows.

E.7. **Converse.** We have shown that (6) is necessary for incentivizability. All that remains is to show that it is sufficient.

Notice that if $\tau_i > 0$ for all i, the product problem is equivalent (in terms of u-optimal choices) to a problem with payoffs $u(a;\theta) = \sum_i \tau_i u_i(a_i, \theta_i)$. In this latter problem, any X satisfying (6) is aligned with u, and is therefore incentivizable. This in turn implies that X is incentivizable in the original problem.

If $\tau_i \leq 0$ for some i, note that, by the proof of Proposition 1, to show that X of the form described in (6) is incentivizable, it suffices to show that $X(a;\theta) =$

$d(\theta) + \sum_i \tau_i u_i(a_i, \theta_i)$ is, where we may assume $|\tau_i| < 1$ for all i. Letting

$$V(r, a, \theta) = \int_0^r X(a; \theta) dx + \int_r^1 x dx - \frac{1}{2} + \sum_i u_i(a_i, \theta_i)$$

$$= rd(\theta) + \sum_i (1 + r\tau_i) u_i(a_i, \theta_i) - \frac{r^2}{2},$$

the V-optimal choice of a is the same as the u-optimal one since $1 + r\tau_i > 0$ for all i and r, and the optimal choice of r is $\mathbb{E}_p[X(a; \theta)]$, as needed.

References

Abdellaoui, Mohammed, Han Bleichrodt, and Cédric Gutierrez (2024). "Unpacking overconfident behavior when betting on oneself". *Management Science* 70.10, pp. 7042–7061.

Amelio, Andrea (2022). "Cognitive uncertainty and overconfidence". ECONtribute Discussion Paper No. 173.

Aoyagi, Masaki, Guillaume R. Fréchette, and Sevgi Yuksel (Dec. 2024). "Beliefs in Repeated Games: An Experiment". *American Economic Review* 114.12, pp. 3944–75.

Arts, Sara, Qiyan Ong, and Jianying Qiu (2024). "Measuring decision confidence". *Experimental Economics*, pp. 1–22.

Ashraf, Nava, James Berry, and Jesse M. Shapiro (Dec. 2010). "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia". *American Economic Review* 100.5, pp. 2383–2413.

Azrieli, Yaron, Christopher P Chambers, and Paul J. Healy (2018). "Incentives in experiments: A theoretical analysis". *Journal of Political Economy* 126.4, pp. 1472–1503.

Bellemare, Charles, Sabine Kröger, and Arthur Van Soest (2008). "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities". *Econometrica* 76.4, pp. 815–839.

Belzil, Christian, Arnaud Maurel, and Modibo Sidibé (2021). "Estimating the Value of Higher Education Financial Aid: Evidence from a Field Experiment". *Journal of Labor Economics* 39.2, pp. 361–395.

Benoît, Jean-Pierre, Juan Dubra, and Don A Moore (2015). "Does the better-than-average effect show that people are overconfident?: Two experiments". *Journal of the European Economic Association* 13.2, pp. 293–329.

Blanco, Mariana, Dirk Engelmann, Alexander K Koch, and Hans-Theo Normann (2010). "Belief elicitation in experiments: Is there a hedging problem?" *Experimental Economics* 13, pp. 412–438.

Blavatskyy, Pavlo R. (Feb. 1, 2009). "Betting on own knowledge: Experimental test of overconfidence". *Journal of Risk and Uncertainty* 38.1, pp. 39–49.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2019). "Beliefs about gender". *American Economic Review* 109.3, pp. 739–773.

Burks, Stephen V, Jeffrey P Carpenter, Lorenz Goette, and Aldo Rustichini (2013). "Overconfidence and social signalling". *Review of Economic Studies* 80.3, pp. 949–983.

Butler, David J and Graham C Loomes (2007). "Imprecision as an account of the preference reversal phenomenon". *American Economic Review* 97.1, pp. 277–297.

Chambers, Christopher P and Nicolas Lambert (2021). "Dynamic belief elicitation". *Econometrica* 89.1, pp. 375–414.

Charness, Gary, Uri Gneezy, and Brianna Halladay (2016). "Experimental methods: Pay one or pay all". *Journal of Economic Behavior & Organization* 131, pp. 141–150.

Charness, Gary, Uri Gneezy, and Vlastimil Rasocha (2021). "Experimental methods: Eliciting beliefs". *Journal of Economic Behavior & Organization* 189, pp. 234–256.

Chassang, Sylvain, Gerard Padró I Miquel, and Erik Snowberg (June 2012). "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments". *American Economic Review* 102.4, pp. 1279–1309.

Clark, Jeremy and Lana Friesen (2009). "Overconfidence in forecasts of own performance: An experimental study". *The Economic Journal* 119.534, pp. 229–251.

Clippel, Geoffroy de, Paola Moscariello, Pietro Ortoleva, and Kareen Rozen (2024). "Caution in the face of complexity". Working paper.

Coffman, Katherine Baldiga (2014). "Evidence on self-stereotyping and the contribution of ideas". *Quarterly Journal of Economics* 129.4, pp. 1625–1660.

Danz, David, Lise Vesterlund, and Alistair J Wilson (2022). "Belief elicitation and behavioral incentive compatibility". *American Economic Review* 112.9, pp. 2851–2883.

Dizon-Ross, Rebecca (Aug. 2019). "Parents' Beliefs about Their Children's Academic Ability: Implications for Educational Investments". *American Economic Review* 109.8, pp. 2728–65.

Dupas, Pascaline (2014). "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment". *Econometrica* 82.1, pp. 197–228.

Eil, David and Justin M. Rao (2011). "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself". *American Economic Journal: Microeconomics* 3.2, pp. 114–138.

Enke, Benjamin and Thomas Graeber (2023). "Cognitive uncertainty". *Quarterly Journal of Economics* 138.4, pp. 2021–2067.

Ertac, Seda (2011). "Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback". *Journal of Economic Behavior & Organization* 80.3, pp. 532–545.

Exley, Christine L. and Kirby Nielsen (Mar. 2024). "The Gender Gap in Confidence: Expected but Not Accounted For". *American Economic Review* 114.3, pp. 851–85.

Fehr, Ernst, Michael Powell, and Tom Wilkening (Apr. 2021). "Behavioral Constraints on the Design of Subgame-Perfect Implementation Mechanisms". *American Economic Review* 111.4, pp. 1055–91.

Gillen, Ben, Erik Snowberg, and Leeat Yariv (2019). "Experimenting with measurement error: Techniques with applications to the Caltech cohort study". *Journal of Political Economy* 127.4, pp. 1826–1863.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (2023). "Designing information provision experiments". *Journal of Economic Literature* 61.1, pp. 3–40.

Healy, Paul J. and Greg Leo (2024). "Belief elicitation: A user's guide". To appear in the Handbook of Experimental Methodology, ed. L. Yariv and E. Snowberg.

Hoff, Karla and Priyanka Pandey (2006). "Discrimination, social identity, and durable inequalities". *American Economic Review* 96.2, pp. 206–211.

Hoffrage, Ulrich (2016). "Overconfidence". *Cognitive Illusions.* Psychology Press, pp. 291–314.

Hossain, Tanjim and Ryo Okui (2013). "The binarized scoring rule". *Review of Economic Studies* 80.3, pp. 984–1001.

Hu, En Hua (2023). "Confidence in Inference". Working paper.

Lambert, Nicolas (2019). "Elicitation and evaluation of statistical forecasts". Working paper.

Lambert, Nicolas, David M Pennock, and Yoav Shoham (2008). "Eliciting properties of probability distributions". *Proceedings of the 9th ACM Conference on Electronic Commerce*, pp. 129–138.

Liberman, Varda (2004). "Local and global judgments of confidence." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30.3, p. 729.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat (2022). "Managing self-confidence: theory and experimental evidence". *Management Science* 68.11, pp. 7793–7817.

Moore, Don A and Paul J Healy (2008). "The trouble with overconfidence." *Psychological Review* 115.2, p. 502.

Morris, Stephen and Takashi Ui (2004). "Best response equivalence". *Games and Economic Behavior* 49.2, pp. 260–287.

Niederle, Muriel and Lise Vesterlund (2007). "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics* 122.3, pp. 1067–1101.

Nyarko, Yaw and Andrew Schotter (2002). "An Experimental Study of Belief Learning Using Elicited Beliefs". *Econometrica* 70.3, pp. 971–1005.

Ortoleva, Pietro and Erik Snowberg (2015). "Overconfidence in political behavior". *American Economic Review* 105.2, pp. 504–535.

Rutström, E. Elisabet and Nathaniel T. Wilcox (2009). "Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test". *Games and Economic Behavior* 67.2, pp. 616–632.

Schlag, Karl H, James Tremewan, and Joël J Van der Weele (2015). "A penny for your thoughts: A survey of methods for eliciting beliefs". *Experimental Economics* 18, pp. 457–490.

Schotter, Andrew and Isabel Trevino (2014). "Belief Elicitation in the Laboratory". *Annual Review of Economics* 6, pp. 103–128.

Serra-Garcia, Marta and Uri Gneezy (Oct. 2021). "Mistakes, Overconfidence, and the Effect of Sharing on Detecting Lies". *American Economic Review* 111.10, pp. 3160–83.

Smith, Cedric AB (1961). "Consistency in statistical inference and decision". *Journal of the Royal Statistical Society: Series B (Methodological)* 23.1, pp. 1–25.

Trautmann, Stefan T and Gijs van de Kuilen (2015). "Belief elicitation: A horse race among truth serums". *The Economic Journal* 125.589, pp. 2116–2135.

Xiang, Yang, Thomas Graeber, Benjamin Enke, and Samuel J Gershman (2021). "Confidence and central tendency in perceptual judgment". *Attention, Perception, & Psychophysics* 83, pp. 3024–3034.

Zimmermann, Florian (2020). "The Dynamics of Motivated Beliefs". *American Economic Review* 110.2, pp. 337–363.