Narrow Inference and Incentive Design*

Alexander Clyde†

March 21, 2025

[CLICK HERE FOR LATEST VERSION]

Abstract

There is evidence that people struggle to do causal inference in complex multidimensional environments. This paper explores the consequences of this in a principal-agent setting. A principal chooses a mechanism to screen an agent. The agent makes choices on multiple dimensions, and infers the effect of each action separately without properly controlling for the other actions. I fully characterize the principal's optimal mechanism when facing an agent who does such 'narrow' inference, and contrast it with their optimal mechanism when the agent is fully rational. I demonstrate when the principal can exploit narrow inference and in what cases they lose out.

Keywords: Behavioural Mechanism Design, Screening, Narrow Bracketing, Misspecified Models.

JEL Classification: D90, D02, D82

*I am grateful to my advisors Ran Spiegler and Philippe Jehiel, for their support and guidance at all stages of this project. I also thank Martin Cripps, Nathan Hancart, Deniz Kattwinkel, Vasiliki Skreta, Duarte Gonçalves, Michael Thaler, Mark Armstrong, Nikita Roketskiy, Elena Ashtari Tafti and Dajana Xhani for helpful comments. I gratefully acknowledge financial support from an Economic and Social Research Council studentship (1923039).

†Department of Economics, University College London; Address: Drayton House, 30 Gordon Street, London, WC1H 0AX; Email: alexander.clyde.econ@gmail.com or alex.clyde@ucl.ac.uk

1 Introduction

Understanding the incentives we face requires understanding how the many choices we make affect outcomes we care about. For example, in the labour market workers have to form beliefs about how their choices of effort, occupation and education affect their final wage. This can require making sophisticated causal inference from any available data. Economic models often assume that people form these beliefs and makes these choices jointly in one over-arching decision problem. However, work in experimental economics suggests that people both fail to consider their choices jointly and struggle to correctly infer the effects of their actions.[1] In line with this, I formulate a model of bounded rationality in which people form beliefs separately in a piecewise way for each decision they face. I call this model of belief formation 'narrow inference'.

Taking into account narrow inference matters for how we should design incentives. Consider a firm that is determining their wage structure. A worker has to decide whether to make human capital investments in technical skills and managerial skills. Assume that the workers beliefs about how these actions affect future earnings have to be consistent with actual data about earnings within the firm. A worker who makes narrow inference forms beliefs about the effect of technical skills and managerial skills on earnings separately. In forming beliefs about the effect of any individual action in this way, they fail to control for other dimensions of action. This leads to a confounding bias that distorts the worker's perception of the size of earnings benefits of obtaining different forms of human capital. The extent of this misperception affects how the firm wants to design their wage structure.

In this paper I analyze such incentive design problems. I consider a principal-agent screening model with a principal who has full understanding of their problem, but an agent who only performs narrow inference. I explore how a principal would design an incentive mechanism if they knew the agent had this form of bounded rationality. I obtain a result characterizing the principal's optimal mech-

[1] In the literature review I discuss work on 'narrow bracketing' from behavioural economics, and work in experimental economics on correlation neglect and causal misperceptions.

anism with an agent who makes narrow inference as the solution to a zero-sum game. I use this result to demonstrate in what cases the principal benefits from the agent making narrow rather than fully rational inference, and in what cases the principal does not. I then explore what happens when the number of dimensions of action grows large. In doing so, I obtain a result that demonstrates the effect of narrow inference on agent's perception of incentives in an optimal mechanism can be quantitatively large.

In the screening model, the agent faces a binary decision problem on whether to take an action or not on multiple dimensions. The principal chooses a function mapping the agent's actions to a transfer, and the agent needs to infer how their actions affect the transfer. There is a large population of agents who differ according to a single dimensional type variable that affects the predictable costs and benefits of the actions. The principal's choice of transfer function screens different types of agents into choosing different actions. In the absence of bounded rationality this problem is standard. The principal and the agent derive the opposite utility from the transfer, and also derive a potentially different predictable utility from the agent's actions that is not mediated by the transfer. The predictable utility of the actions is strictly increasing in the type for the agent and additively separable across dimensions for both the agent and the principal, but the principal can choose a transfer function where there is an interactive effect.

An agent who makes narrow inference calculates the effect of each dimension's action on the transfer separately. Their beliefs about the effect of an action on a given dimension must be consistent with data on the population level average transfer conditional on that action. The difference between the average population level transfers between any two actions in a given dimension is then used to estimate the relative effect of each action on the transfer. This is a naive way to estimate the 'treatment effect' of any action. It can lead to distorted beliefs if the distribution over actions are correlated across dimensions, something that is possible due to joint dependence on the type variable. The estimated effect of the action on the transfer is then biased from confounding. This holds even when the true transfer function is additively separable, as the inferential failure of the agent involves neglecting the correlation in the data and not just in misspecifiying the functional form of the transfer function.

In what follows, I develop the firm-worker example to illustrate narrow inference and to demonstrate how the principal might benefit from agents using narrow inference.

Example 1. The agent has two dimensions of action, whether to obtain technical skills $a_1 = 1$ or not $a_1 = 0$ and whether obtain managerial skills $a_2 = 1$ or not $a_2 = 0$. The principal, a firm, sets earnings schedule for roles within the firm that depends jointly on these two actions $t : A_1 \times A_2 \to \mathbb{R}$. There are three types of agent $s \in \{0, 1, 2\}$. The probabilities of the types are denoted $p_0,\ p_1,\ p_2 \in (0, 1)$ respectively. The agent's utility depends on their type s, their actions a_1, a_2 and the transfer t.

$$t(a_1, a_2) - (3 - s)(a_1 + a_2)$$

Where $-(3 - s)(a_1 + a_2)$ is the predictable utility cost of making human capital investments for the worker. Suppose that the principal wants to implement that the type $s = 0$ chooses neither action, the type $s = 1$ obtains technical skills $a_1 = 1$ but not managerial skills $a_2 = 0$, while the highest type $s = 2$ obtains both $a_1 = a_2 = 1$. An earnings function that ensures that rational agents will act this way must satisfy the following incentive constraints. The first ensures that type $s = 1$ chooses $(1, 0)$ over $(0, 0)$ and the second ensures type $s = 2$ chooses action $(1, 1)$ over $(1, 0)$.

$$t(1, 0) - 2 \geq t(0, 0)$$
$$t(1, 1) - 2 \geq t(1, 0) - 1$$

Choosing t such that these incentive constraints bind allows the principal to minimize the earnings paid to types $s = 1$ and $s = 2$. This means $t(1, 1) > t(1, 0) > t(0, 0)$. These local incentive constraints binding suffices for all incentive constraints to hold.

Now consider what happens if the agent uses narrow inference, but the transfer function is fixed at those that bind for rational incentive compatibility. They expect the earnings from any action to be the average population level earnings of those who have taken that action. For an agent of type $s = 1$ making narrow inference to want to invest in technical skills requires that

$$\frac{p_1}{p_1 + p_2} t(1,0) + \frac{p_2}{p_1 + p_2} t(1,1) - 2 \geq t(0,0)$$

Their narrow beliefs about the earnings from acquiring technical skills pool the earnings of both the types who obtain these skills, who have proportion $p_1 + p_2$ in the population of types. The narrow beliefs about the earnings from not acquiring technical skills are in line with rational beliefs, since only the lowest type $s = 0$ doesn't acquire these skills. Since $t(1,1) > t(1,0)$, the narrow perception of the expected earnings benefit of technical skills is biased upward from the true effect. It fails to adjust for the fact that a type $s = 1$ agent who is on the margin between obtaining technical skills does not obtain managerial skills and thus has lower future earnings than the average worker who obtains technical skills.

Similarly for the type $s = 2$ agent to want to obtain managerial skills under narrow inference requires that

$$t(1,1) - 1 \geq \frac{p_1}{p_1 + p_0} t(1,0) - \frac{p_0}{p_1 + p_0} t(0,0)$$

As $t(1,0) > t(0,0)$, their perception of the earnings from not obtaining managerial skills is biased downwards. The expected earnings for those who do not obtain education mixes the earnings of those who get work experience and those who do not. It is therefore less than the earnings obtained by the types on the margin of obtaining managerial skills, type $s = 2$, all of whom obtain technical skills.

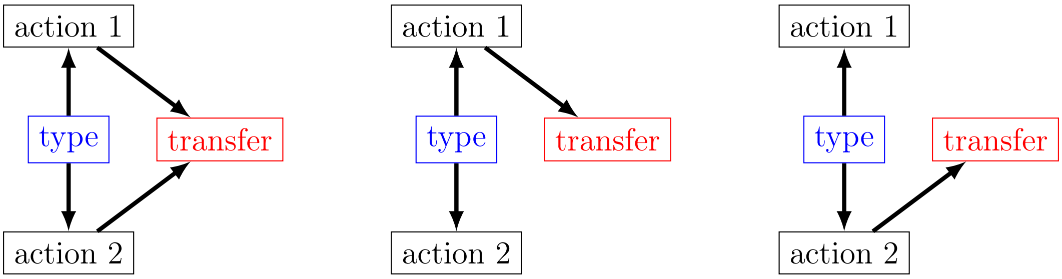
This upward bias in the incentives the agent perceives allows the principal to implement the same action choices for each type while reducing earnings across the type distribution. △

My analysis of the design problem proceeds as follows. First, in order to contrast the principal's optimal mechanism when agents make narrow inference to that with fully rational agents, I first state results describing the principal's optimal mechanism in the rational benchmark. This involves applying results in single-dimension monopolistic screening problems adapted to the multidimensional action setting. Under a standard regularity assumption, the principal's optimal mechanism is fully separable across dimensions. Facing such a mechanism, the agent chooses a strategy that on each dimension selects the action if and only if their type is above a dimension-specific threshold.

I then obtain a result characterizing the principal's optimal mechanism under narrow inference. In this characterization, the principal plays a zero-sum game against an adversarial player who can shrink the agent's predictable utility from the actions on any dimension by some constant factor that lies between one and zero, with the shrinkage factors summing to one across dimensions. The principal's optimal mechanism under narrow inference then solves the same problem as in the rational benchmark except with the shrunk predictable utilities. The result allows us to both solve for the principal's optimal mechanism for specific parameterizations, and also enables us to easily compare how the mechanism differs from that with rational agents.

I show using the characterization result what effect narrow inference has on this threshold strategy and on the principal's welfare relative to the rational benchmark. This involves considering two different cases; one where the agent's actions have predictable utility costs to the agent but benefits to principal on all dimensions, and one where all actions have predictable benefits to the agents but costs to the principal.[2] The first case fits the firm-worker example, where human capital investments have acquisition costs for the worker but increase profitability for the firm. For an example that fits the second case, consider a buyer-seller setting. The principal is a seller of two goods, such as computers and software, for which the

[2]Whether the actions have a positive or negative predictable utility for the agent is not an arbitrary re-labelling of the actions, as we normalize the predictable utility from the zero action to zero. Without the normalization the result would be stated in terms of the difference in predictable utilities between the two actions.



(a) True causal model (b) Dimension 1 beliefs (c) Dimension 2 beliefs

Figure 1: DAG illustrating confounding bias with two dimensions of action.

ultimate price can be opaque to the buyer and can be affected by both purchases in a joint way. The buyer gets positive predictable utility from buying a more advanced computer and from better software. They make narrow inference about what the effect of each choice is on the ultimate price they are paying for their computing.

In the case where actions predictably cost the agent but benefit the principal, when facing the principal's optimal mechanism the agent's thresholds are lower under narrow inference than when they are rational and the principal is always weakly better-off under narrow inference. Lower thresholds mean a greater proportion of types take the action on any dimension. On the other hand, when actions predictably cost the principal but benefit the agent then thresholds are higher and the principal is always weakly worse-off under narrow inference. These effects are the consequence of the agent overestimating the causal effects of their actions on transfers due to confounding bias. Taking the action on one dimension is associated with taking the action on others, as both are chosen by higher type agents, and when agents neglect this their beliefs of about the effect of the action are biased upwards. Figure 1 illustrates the confounding bias from omitting the actions on other dimensions using the Directed Acyclic Graph (DAG) notation of Pearl (2009). I show that the principal's gains and losses from narrow inference are purely due to this confounding, as under narrow inference the principal's optimal mechanism has a separable transfer function just as in the rational case.

The effect of narrow inference can be quantitatively large. Assume that the predictable utility of the actions is identical across dimensions for both the principal and agent. I consider what happens if total predictable utilities are split over actions dimensions, so that we have a sequence of problems with a larger number of action dimensions but the same total predictable utilities from the actions for the agent and the principal. If the number of dimensions is big enough then in the principal's optimal mechanism under narrow inference either all types of agents take the actions on all dimensions, or no agents take the actions on any dimension. In the first case, the principal is able to implement this strategy at arbitrarily small cost in terms of transfers.

Finally, I explore extensions that vary the form of the participation constraint and consider the consequences of dropping the regularity assumption. In the main model, I assume the agent also makes narrow dimension-by-dimension decisions on whether to participate in the mechanism. This is motivated by an interpretation that non-participation involves resorting to an outside option where the agent takes no action and obtains a default transfer. The agent treats the non-participation decision as they treat decisions within the mechanism. They believe that they can use the outside option narrowly, taking it on some dimensions but continuing to participate in the mechanism on others. I consider a version of the model where the agent instead perceives the participation decision as a joint one, taking the sum of perceived narrow utility across dimensions as the value of participation. I demonstrate how we can modify the characterization result under this alternative participation constraint. The results on the principal's welfare and on large dimensionality still hold.

Literature Review

My paper builds on several distinct but related strands of literature. Experimental work in psychology and economics documents that people make inferential errors similar to narrow inference. Enke and Zimmermann (2019) find subjects fail to adjust for correlation between multiple information sources. Similar logic extends to predictive tasks, in He and Kučinskas (2024) subjects' forecasting performance deteriorates when information from a single variable is split into two. Fernbach et al. (2010) present evidence suggesting people focus narrowly on a few variables when trying to make causal predictions. In line with this, Graeber (2023) finds subjects ignore the effect of variables that are not directly involved in a predictive task despite these variables containing valuable information.

Narrow inference involves causal misperceptions, but also thinking about decision problems narrowly. The literature on narrow bracketing considers decision makers who break decision problems into smaller sub-problems without accounting for how these decisions interact in the larger joint problem. Work in this area; Tversky and Kahneman (1981), Thaler (1985), Thaler (1999), Read et al. (1999), Rabin and Weizsäcker (2009), has both documented evidence for and explored the theoretical implications of narrow decision making. Recent work exploring theoretical foundations for narrow behaviour includes Kőszegi and Matějka (2020), who use a model of costly information acquisition to explain both mental accounting and naive diversification. Lian (2021) builds a theory of 'narrow thinking', which models decision makers as playing an incomplete information bayesian game between multiple-selves. Camara (2022) shows that computability constraints imply a form of narrow choice bracketing. Vorjohann (2024) presents an axiomatization for narrow bracketing that can resolve the tension between narrow decision making and the fact globally choice must satisfy a budget balance condition.

In modelling agents with narrow causal perceptions, this paper builds on work studying decision making by agents using misspecified models of how action choices map into consequences. There is a growing literature on the Berk-Nash Equilibrium of Esponda and Pouzo (2016), a solution concept founded as the limit of a process of misspecified learning; (Heidhues et al., 2018; Frick et al., 2020; Bohren and Hauser, 2021; Fudenberg et al., 2021). Another connected literature is that on modelling causal misperceptions using Bayesian Networks; (Spiegler, 2016; Eliaz and Spiegler, 2020). Schumacher and Thysen (2022) use this Bayesian Network approach in a principal-agent moral hazard problem where the agent has causal misperceptions of how their actions map into output. In Eliaz and Spiegler (2024) a Bayesian Network formalism is used to model the design of narratives for misspecified news consumers by media organizations.

Earlier work on design when agents misperceive incentives by Rubinstein (1993) and Piccione and Rubinstein (2003) explores monopolistic pricing when customers have a coarse misperception of any pricing strategy. Eyster and Rabin (2005) consider a solution concept for games where players neglect correlation between information and opposing players actions, and apply their concept to bilateral trade and auction settings. Jehiel (2005) develops an equilibrium concept for extensive form games —Analogy Based Expectation Equilibrium (ABEE)— in which players have coarse misperceptions of other players' strategies. The behaviour of the principal and the agent under narrow inference can be formulated as an ABEE of an extensive form game, and I discuss this in more detail in Section 5.3. The first papers to explicitly apply the ABEE concept to design problems are Jehiel (2011) and Jehiel and Mierendorff (2024). In Jehiel (2011), an auction designer manipulates bidders who do not perceive how the distribution of bids varies with different auction formats and the identities of different bidders. Similarly, in Jehiel and Mierendorff (2024) a proportion of bidders form beliefs about how signals of their own valuation of an item vary with opponents bids in a way that neglects correlation between their own signal and the signals of the other bidders. A paper explicitly relating narrow bracketing and auction design is Eisenhuth (2019), who considers a model where loss-averse bidders in an auction narrowly bracket gains and losses relative to a reference point, and studies the implications for auction design.

In contributing to the small literature on mechanism design where agents use misspecified models, this paper also contributes to a larger literature on mechanism design that takes into account agents' limited rationality in a variety of other dimensions. A detailed review can be found in Kőszegi (2014). This includes work in contract theory (Eliaz and Spiegler, 2006), (Heidhues and Kőszegi, 2010), (Herweg et al., 2010) and optimal taxation (O'Donoghue and Rabin, 2006), (Spinnewijn, 2015), (Farhi and Gabaix, 2020), (Lockwood, 2020). Some models of optimal income taxation under behavioural biases are built on empirical findings that people have coarse misperceptions of income tax schedules that are similar to narrow inference; (Liebman and Zeckhauser, 2004), (Feldman et al., 2016), (Rees-Jones and Taubinsky, 2020).

2 Model

An agent faces a multidimensional decision. Let $A = \{0,1\}^n$ be the agent's set of feasible action profiles. I refer to $i \in \{1,..,n\} \equiv N$ as a dimension, such that a_i is the agent's action in dimension i. The agent has a type that lies in a bounded interval $s \in S \equiv [0,1]$. This type is drawn from an atomless distribution that admits a density $p(s)$ such that $p(s) > 0$ for all $s \in S$. Denote the cdf of the distribution by $P(s) = \int_0^s p(\tilde{s})d\tilde{s}$.

The dimension i action a_i generates a predictable utility $v_i(s)a_i$, where $v_i(s)$ is strictly increasing, continuously differentiable in s and can be positive or negative. In addition to the predictable utility, the agent receives utility from a transfer t that needs to be inferred. The utility of an agent of type s, choosing action a with transfer $t \in \mathbb{R}$ is

$$u(s,a,t) = \sum_{i \in N} v_i(s)a_i + t \tag{1}$$

The principal derives benefit/costs $w_i a_i$ from an action a_i, where $w_i \in \mathbb{R}$. The transfer $t \in \mathbb{R}$ represents a zero-sum transfer of surplus between the agent and the principal. The principal's payoff given actions a and transfer $t \in \mathbb{R}$ is

$$W(a,t) = -t + \sum_{i \in N} w_i a_i \tag{2}$$

Throughout the paper, I make the following standard regularity assumption on the type distribution. In Section 5.2, I explore the implications of dropping this assumption.

Assumption 1. For every dimension $i \in N$

$$\phi_i(s) = v_i(s) - \frac{1 - P(s)}{p(s)} v_i'(s)$$

is strictly increasing in $s \in S$. We refer to this property as increasing virtual values (IVV).

2.1 Mechanisms

I focus on a natural class of indirect mechanisms. Before the agent takes any actions the principal commits to a mechanism, which consists of a function mapping actions to transfers $t : A \to \mathbb{R}$. After learning their type, the agent chooses a distribution over actions according to a strategy $g : S \to \Delta(A)$. The marginal distribution over actions in dimension i is denoted by $g_i(a_i|s) = \sum_{a_{-i} \in A_{-i}} g(a_i, a_{-i}|s)$.

For a transfer function $t \in \mathbb{R}^A$, given a strategy g the expected payoff for the principal is

$$W(t, g) = \int_0^1 \sum_{a \in A} [-t(a) + \sum_{i \in N} w_i a_i] g(a|s)\, p(s) ds \tag{3}$$

The restriction to this class of mechanisms is simple to reconcile with narrow inference. Under narrow inference, the agent perceives the transfer as measurable only with respect to their own actions. Suppose the principal could choose a more general mechanism in which the transfer function varied with an arbitrary message space as well as the actions. The principal could present information on how the transfer varies with more finely grained messages, drawing the agent's attention to the joint multidimensional nature of their problem and undoing the narrow inference.

In Section 3, I show the restriction makes no difference to the analysis of the principal's optimal mechanism in the rational case. Under the optimal mechanism the agent chooses a strategy that is deterministic, and as such can be implemented with a transfer function that only depends on the chosen action.

2.2 Model Interpretations

The model allows actions to have both a positive and negative effect on payoffs. The sign of $v_i(s)$ determines whether a type s agent has predictable positive utility from action $a_i = 1$ or predictable disutility. Likewise, the direct effect of actions on the principal's payoff can be positive ($w_i \geq 0$) or negative ($w_i \leq 0$).

The transfer function t can be interpreted as the division of the surplus or costs from actions. The surplus or costs of the actions a is given by $\sum_{i \in N} w_i a_i$. The transfer function is then the share of that surplus the agents receive, or the share of the costs they must bear.

This framework can capture the stories given in the introduction. Suppose the agent is a buyer of computers and software and the principal a seller. In this story, $v_i(s) > 0$ represents the predictable payoff benefit to the type s firm of purchasing the good $a_i = 1$, while $w_i < 0$ represents the cost to the firm of production. The seller then chooses a pricing schedule t.

In another story the agent is a worker and the principal a firm. Here $v_i(s) < 0$ is the cost of obtaining human capital and $w_i > 0$ is the benefit to the firm of the human capital the worker acquires. The firm then chooses a earnings schedule t, which depends on the skills the workers acquire.

2.3 Rational Inference

Given a strategy g, write the expected utility of an agent of type s as

$$U(s) = \sum_{a \in A} g(a|s) u(s, a, t(a)) = \sum_{a \in A} g(a|s) [\sum_{i \in N} v_i(s) a_i + t(a)] \tag{4}$$

Incentive Compatibility (IC) of strategy g under transfer function $t \in \mathbb{R}^A$ requires that g is a best response to t. This means for any $s \in S$, $a \in supp(g(.|s))$ and any $a' \in A$

$$\sum_{i \in N} v_i(s) a_i + t(a) \geq \sum_{i \in N} v_i(s) a'_i + t(a') \tag{5}$$

The agent always has the option of not participating in the mechanism, taking the actions $a = 0$ and obtaining zero transfer. We then have the following participation constraint; for all $s \in S$

$$U(s) \geq 0 \tag{6}$$

2.4 Narrow Inference

Given a strategy g, an unconditional distribution over actions in A is induced as follows.

$$g(a) = \int_0^1 g(a|s)\, p(s) ds \tag{7}$$

Let the marginal over an action in dimension i be denoted $g_i(a_i) = \sum_{a_{-i} \in A_{-i}} g(a_i, a_{-i})$. I use the terms action distribution and strategy interchangeably throughout the paper.

The agent forms narrow perceptions of the mechanism's transfer function. In particular an agent believes when taking a decision in dimension i that in expectation they will receive $\overline{t}_i(a_i)$ if they take action a_i. When $g_i(a_i) > 0$ we require that this expectation is consistent with the actual conditional expectation of transfers given a_i.

$$\overline{t}_i(a_i) = \sum_{a_{-i} \in A_{-i}} \frac{g(a_i, a_{-i})}{g_i(a_i)} t(a_i, a_{-i}) \tag{8}$$

Denote the vectors of beliefs $\overline{t}(a) = (\overline{t}_i(a_i))_{i=1}^n \in \mathbb{R}^n$ and $\overline{t} = (\overline{t}(a))_{a \in A} \in \mathbb{R}^{2n}$. When $g_i(a_i) = 0$, we allow $\overline{t}_i(a_i)$ to take on arbitrary values. Analogously to the rational benchmark, these 'off-path' actions do not affect the principal's objective and $\overline{t}$ can be set to ensure incentive compatibility. Henceforth, I refer to agents performing narrow inference as 'narrow agents'.

A narrow agent imposes an additively separable form on their estimate of the transfer function using $\overline{t}$. This gives them the following perceived expected utility from strategy g when they are of type $s \in S$.

$$\overline{U}(s) = \sum_{i \in N} \sum_{a_i \in A_i} g_i(a_i|s)[v_i(s)a_i + \overline{t}_i(a_i)] = \sum_{i \in N} \overline{U}_i(s) \tag{9}$$

Where

$$\overline{U}_i(s) = \sum_{a_i \in A_i} g_i(a_i|s)[v_i(s)a_i + \overline{t}_i(a_i)] \tag{10}$$

denotes the narrow perceived expected utility of type s in dimension i. A strategy g is *narrow incentive compatible* (NIC) if for any dimension $i \in N$, type $s \in S$ and actions $a_i \in supp(g_i(.|s))$, $a_i' \in A_i$

$$v_i(s)a_i + \overline{t}_i(a_i) \geq v_i(s)a_i' + \overline{t}_i(a_i') \tag{11}$$

We assume there are dimension by dimension *narrow participation constraints*; for all $s \in S$, $i \in N$

$$\overline{U}_i(s) \geq 0 \tag{12}$$

This fits the following interpretation; the agent believes they can reject any additional effect on the transfer resulting from participation in the mechanism separately on each dimension, whilst still obtaining the effect from participation on other dimensions. This is in line with the agent believing the true transfer function is additive. Although this appears to add participation constraints relative to the rational benchmark, in practice it does not. In Section 3, I show that the transfer function in the principal's optimal mechanism with a rational agent is additively separable and thus also satisfies these dimension-by-dimension constraints. I consider an alternative joint participation constraint in Section 5.1.

3 Rational Benchmark

With rational agents, we have a screening problem with a single dimension of type but multiple dimensions of action. I restate existing results adapted to our setting.[3]

[3] In particular Proposition 3.1 of Carroll (2017). Earlier work by Mirman and Sibley (1980) analyzes the monopoly screening model with many dimensions of action but a single dimension of type, but assumes directly that the IC action on any dimension is increasing in the type. In Carroll (2017) this is explicitly shown to hold under the regularity (IVV) assumption.

It will be shown that the agent's strategy under the principal's optimal mechanism with rational agents takes a *threshold form* where there is a potentially different threshold $\hat{s}_i \in S$ on each dimension such that $g_i(1|s) = \mathbb{1}\{s \geq \hat{s}_i\}$. Let the vector of thresholds across dimensions be denoted $\hat{s} = (\hat{s}_i)_{i \in N} \in [0,1]^n$. We can characterize the principal's problem in terms of choosing these thresholds. Denote the value of the principal's objective under threshold strategy $\hat{s}$ by $W(\hat{s})$.

Proposition 1. *Assume the IVV assumption holds. The principal maximizes their objective over all IC mechanisms that satisfy the participation constraint if and only if they choose a transfer function implementing a threshold strategy that solves the following problem.*

$$\max_{\hat{s} \in [0,1]^n} W(\hat{s}) = \sum_{i \in N} \int_{\hat{s}_i}^{1} (\phi_i(s) + w_i)\, p(s)\, ds \tag{13}$$

The principal's value under an objective maximizing mechanism can be achieved by an additively separable transfer function

$$t(a_1, ..., a_n) = \sum_{i \in N} t^i(a_i) \tag{14}$$

$$t^i(0) = 0, t^i(1) = -v_i(\hat{s}_i) \text{ for all } i \in N \tag{15}$$

Proof. In Appendix □

Thus, the principal's optimal transfer function can be treated as the sum of separate transfer functions, one for each dimension. This does not result directly from IC, but rather from the optimality for the principal of implementing a threshold strategy when IVV holds. We will see in Section 5.2 that without IVV it can be optimal for the principal to choose a non-separable transfer function, and that under such a transfer function a non-threshold strategy is implemented and thus IC.

4 Narrow Agents

The solution to the principal's design problem with narrow agents can be characterized as a zero-sum game between the principal and an adversarial player. In this game, the principal faces the same design problem as in the rational benchmark except the predictable utilities are scaled by some factor $\beta_i \in [0, 1]$ in each dimension. The predictable utility in dimension i is then $\beta_i v_i(s)$, and the scaling factors sum to one across dimensions $\sum_{i \in N} \beta_i = 1$. The principal chooses a mechanism to maximize their objective while the adversarial player simultaneously chooses the scaling factors to minimize the value of the objective. The result shows that the agent's strategy and value of the principal's objective under the principal's optimal mechanism with narrow agents coincide with those that arise as the solution to this zero sum game with rational agents.

An interpretation of the shrinkage factors is as follows. For a given perceived size of incentives, you have a transfer function that would achieve these perceived incentives under narrow inference, and a transfer function that would achieve these perceived incentives under the rational benchmark. The shrinkage factors give how much these transfers for the rational benchmark would have to be scaled to achieve the same expected transfer as under narrow inference. The fact the shrinkage factors must be between zero and one and sum to one demonstrates that narrow inference amplifies the size of incentives relative to the rational benchmark.

4.1 Main Characterization Result

Theorem 1. *Assume the IVV assumption holds. The principal maximizes their objective over all NIC mechanisms that satisfy the narrow participation constraints if and only if they choose a transfer function that implements a threshold strategy that solves*

$$\min_{\beta \in [0,1]^n : \sum_{i \in N} \beta_i = 1} \max_{\hat{s} \in [0,1]^n} \overline{W}(\hat{s}; \beta) = \max_{\hat{s} \in [0,1]^n} \min_{\beta \in [0,1]^n : \sum_{i \in N} \beta_i = 1} \overline{W}(\hat{s}; \beta) \tag{16}$$

with the value of the principal's objective given by

$$\overline{W}(\hat{s};\beta) = \sum_{i\in N} \int_{\hat{s}_i}^{1} (\beta_i \phi_i(s) + w_i)\, p(s)\, ds \tag{17}$$

Proof. In Appendix □

To see some intuition for this result, consider the case with two dimensions and symmetric predictable utility $v_i(s) = v(s)$ and principal's direct utility $w_i = w$ from actions across all dimensions $i \in N$. With a rational agent, from Proposition 1 the principal's optimum sets a threshold $\hat{s}$ so that the action is taken on all dimensions for all types above and the zero action is taken on all dimensions for all types below. This optimum is induced with a transfer function that is additive and identical across dimensions $t(a_1, a_2) = t^1(a_1) + t^2(a_2)$ and $t^1(1) = t^2(1) = \tilde{t}(1)$, $t^1(0) = t^2(0) = 0$. With a narrow agent, under the same strategy and transfer function the agents double count the effect of each action on the transfer. In each dimension, they believe that the transfer resulting from taking the action $a_i = 1$ is $2 \cdot \tilde{t}(1)$ and the transfer resulting from $a_i = 0$ is 0. This double-counting is the result of confounding neglect; the agent fails to adjust for the fact that every type who takes action $a_1 = 1$ also takes action $a_2 = 1$. The principal then has to half the size of the the difference in transfers in order to maintain the same thresholds $\frac{1}{2}\tilde{t}(1)$. This has the same effect as scaling the predictable utilities down by $\frac{1}{2}$ in each dimension when the agent is rational.

The result extends this logic to asymmetric cases. It allows us to both solve for the principal's optimal mechanism with narrow agents and also demonstrates the connection between any given problem with narrow agents to the rational benchmark. I use the characterization to obtain additional results. I give conditions under which the principal does and does not benefit from facing narrow over rational agents, and how the implemented strategy changes between the two cases. I then explore the effect of symmetric predictable utility across dimensions and what happens when the number of dimensions grows large. First, I present some preliminaries that are used in obtaining the characterization.

4.2 Preliminaries for Characterization Result

I first obtain a result characterizing how narrow expected utilities relate to the implemented strategy. In particular we obtain a version of the envelope theorem for each dimension separately. I then consider which beliefs can be induced by a transfer function. I show that beliefs must satisfy a statistical correctness constraint, and that for any deterministic threshold strategy there is a valid additively separable transfer function that induces beliefs such that the strategy is NIC.

Since beliefs $\bar{t}$ can only depend on a, given a fixed strategy g we cannot necessarily obtain any value of the narrow expected utility in dimension i; $\overline{U}_i(s)$ from (11). We say that narrow expected utilities $(\overline{U}_i(s))_{s\in S, i\in N}$ can be *achieved* given strategy g if there exists a transfer function that induces beliefs $\bar{t}$ such that for each dimension $i \in N$ and every $s \in S$

$$\sum_{a_i \in A_i} g_i(a_i|s)\bar{t}_i(a_i) = -v_i(s) \sum_{a_i \in A_i} g_i(a_i|s)a_i + \overline{U}_i(s) \tag{18}$$

Lemma 1. *A strategy g and narrow expected utilities $(\overline{U}(s))_{s\in S}$ that can be achieved given g are NIC if and only if*

1. *The strategy is monotonic on each dimension; that is for all $i \in N$ we have that*
$$\sum_{a_i \in A_i} a_i g_i(a_i|s) \tag{19}$$
is increasing in $s \in S$.

2. *On each dimension $i \in N$, $\overline{U}_i(s)$ is increasing in $s \in S$.*

3. *On each dimension $i \in N$, the following envelope condition holds for any two types $s, s' \in S$*
$$\overline{U}_i(s) = \overline{U}_i(s') + \int_{s'}^{s} v_i'(z) \sum_{a_i \in A_i} a_i g_i(a_i|z)dz \tag{20}$$

Proof. In Appendix □

The requirement that an NIC strategy has increasing predictable utility in type separately on each dimension means that any deterministic strategy in an NIC mechanism must have a threshold form. This differs from the rational case where a threshold strategy is optimal for the principal under the IVV assumption, but is not an implication of IC.

It will be useful in characterizing the principal's optimal mechanism under NIC to show how beliefs and the transfer function relate for any fixed distribution over actions. The following result shows when we can write the transfer distribution in terms of the beliefs over the expected transfer in either dimension. It gives a standard statistical correctness result that applies to beliefs formed using Bayesian Networks under perfect Direct Acyclic Graphs (DAGs) (Spiegler, 2020).

Lemma 2 (Statistical Correctness)**.** *Given any distribution over actions g, for any two dimensions $i, j \in N$ we have that beliefs $\bar{t}_i, \bar{t}_j$ satisfy*

$$\sum_{a_i \in A_i} g_i(a_i)\bar{t}_i(a_i) = \sum_{a_j \in A_j} g_j(a_j)\bar{t}_j(a_j) \tag{21}$$

Proof. Rearranging the expected transfer gives us the first part. For any $i \in N$:

$$\sum_{a \in A} g(a)t(a) = \sum_{a_i \in A_i} g_i(a_i) \sum_{a_{-i} \in A_{-i}} \frac{g(a_i, a_{-i})}{g_i(a_i)} t(a_i, a_{-i}) = \sum_{a_i \in A_i} g_i(a_i)\bar{t}_i(a_i)$$

□

This statistical correctness is necessary but not sufficient for a transfer function to exist that induces given beliefs for a fixed strategy and action distribution. For an example of beliefs that satisfy the statistical correctness constraint but cannot be induced, consider the case with $N = \{1, 2\}$ and $g(1,1) = g(0,0) = \frac{1}{2}$. If beliefs do not also satisfy $\bar{t}_1(1) = \bar{t}_2(1)$ and $\bar{t}_1(0) = \bar{t}_2(0)$, then there is no transfer function implementing these beliefs under this action distribution.

The principal's objective can be written in terms of beliefs. This means it is useful to work directly with beliefs rather than the underlying transfer function when we characterize the principal's optimal mechanism. Although the statistical correctness constraint is not sufficient, it will be sufficient if the strategy of the agent takes a deterministic threshold form. A *deterministic threshold strategy* is such that for all $i \in N$ there is a threshold $\hat{s}_i \in [0,1]$ so that the action a_i is only taken by types above that threshold; $g_i(1|s) = \mathbb{1}\{s \geq \hat{s}_i\}$. I now show that any NIC strategy must take a deterministic threshold form.

Lemma 3. *Every NIC strategy takes a* ***deterministic threshold form*** *for almost all $s \in S$.*

Any strategy g that takes takes a ***deterministic threshold form*** *is NIC if there exists a transfer function t that together with g induces beliefs that for all $i \in N$ satisfy*

$$\bar{t}_i(1) = \bar{t}_i(0) - v_i(\hat{s}_i) \tag{22}$$

Proof. Let $\hat{U}_i(s, a_i) = v_i(s)a_i + \bar{t}_i(a_i)$. Clearly $\hat{U}_i(s,1) - \hat{U}_i(s,0) = v_i(s) + \bar{t}_i(1) - \bar{t}_i(0)$ is increasing in s. There is a threshold $\hat{s}_i$ such that $\hat{U}_i(s,1) > \hat{U}_i(s,0)$ for all $s > \hat{s}_i$, in which case type s will choose $a_i = 1$. The same logic applies for types $s < \hat{s}_i$ who will choose $a_i = 0$. The types who are indifferent have measure zero, so it is without loss for them to choose $a_i = 1$ also.

Given a deterministic threshold strategy, the beliefs in the proposition statement ensure that the agent is indifferent between taking either action at the threshold.

$$v_i(\hat{s}_i) + \bar{t}_i(1) = \bar{t}_i(0)$$

□

The next result shows how any deterministic threshold strategy can be made NIC by a transfer function that is additive across dimensions. This means a transfer function exists that implements any beliefs that satisfy statistical correctness, if the agent is choosing actions according to a deterministic threshold strategy.

Proposition 2. *For any deterministic threshold strategy g, we can construct a transfer function t that induces beliefs $\bar{t}$ so that g is NIC. The constructed transfer* *function is additive; for any* $a_{-i}, \tilde{a}_{-i} \in A_{-i}$ *we have that*

$$t(1, \tilde{a}_{-i}) - t(0, \tilde{a}_{-i}) = t(1, a_{-i}) - t(0, a_{-i}) \tag{23}$$

Moreover, any transfer function $\tilde{t}$ *such that* g *is NIC can only differ from this additive* t *at action combinations that do not occur under* g*;* $t(a) \neq \tilde{t}(a)$ *only if* $g(a|s) = 0$ *for all* $s \in S$*.*

Proof. In Appendix □

Proposition 2 means that if the principal implements a deterministic threshold strategy, they have no payoff gain from implementing a transfer function that is not additive. The principal can exploit two features of the narrow agents misperception; that they can only perceive of the transfer function as additive and that their beliefs do not account for confounding bias. A non-deterministic strategy could enable the principal to implement a non-additive transfer function. However, even if the principal could force the agent to randomize, the proof of Theorem 1 shows it is optimal for the principal to implement a deterministic threshold strategy with an additively separable transfer function. Thus, the principal does not exploit the agent's false perception of the transfer function as additive in their choice of optimal mechanism.

The transfer function can be constructed as additive because under a deterministic threshold strategy, there are action combinations that are not chosen by any type of agent. Take the case with two dimensions of action, and assume that the agent's strategy is such that threshold for the first dimension is lower than the threshold in the second dimension; $\hat{s}_1 < \hat{s}_2$. Types in interval $[0, \hat{s}_1)$ choose actions $(0, 0)$, types in $[\hat{s}_1, \hat{s}_2)$ choose $(1, 0)$ and types in $[\hat{s}_2, 1]$ choose $(1, 1)$. The transfer for actions $(0, 1)$, chosen by no type, can be set so that the transfer function is additive. The case where $\hat{s}_1 \geq \hat{s}_2$ is symmetric. Proposition 2 extends this logic to the general case with any number of dimensions. The transfer function is constructed recursively so as to induce the given beliefs.

The proof of Theorem 1 works as follows. First it uses Lemma 1 to write both the principal's objective and the statistical correctness constraint from Lemma 2 only in terms of the strategy and the narrow perceived utility of the type taking action zero on one of the dimensions; $\overline{U}_i(0)$. It then shows that a minimax upper bound to this constrained problem is solved by implementing deterministic threshold strategies on each dimension, with the β weights in the proof coming from a rewriting of the Lagrange multipliers from the statistical correctness constraint. Under IVV, we can apply a standard minimax theorem argument to obtain a saddle point for this upper bound problem. Finally, we can show that we can achieve this minimax upper bound with a transfer function that solves the full problem using Proposition 2.

4.3 Effect of Narrow Agents on the Principal's Welfare

Using the characterization of the principal's optimal mechanism, I show that when actions have predictable costs for all types of the agent then the principal is better off under narrow inference. Conversely, when actions have predictable benefits for all types then the principal is worse off.

Proposition 3. *Assume the IVV assumption holds.*

1. *When for every $i \in N$ we have $v_i(s) \leq 0$ for all $s \in S$, the principal can obtain at least as high an objective value when the agent is narrow compared to the rational benchmark.*

2. *When for every $i \in N$ we have $v_i(s) \geq 0$ for all $s \in S$, the principal obtains at least as high an objective value in the rational benchmark compared to when the agent is narrow.*

Proof. In Appendix □

The intuition for this result is as follows. When the agent's action is costly in terms of the agent's predictable utility, for any threshold strategy the principal makes transfers to the agent. The single dimension of type results in actions in one dimension being positively correlated with actions on any other dimension.

Since actions result in higher transfers to the agent, this leads a narrow agent to overestimate the transfer they will get from the principal.

Rational agents adjust for the fact their type is on the margin between taking an action or not, so know they will receive a lower transfer than the average obtained by agents taking that action. The overestimation of the transfer by narrow agents means less transfer has to be given to higher type agents in order to implement any given strategy. When actions have a predictable utility benefits to the agent, the principal is a net receiver of transfers from the agent. For clearer intuition, consider the application where the principal is a profit-maximizing seller of multiple goods and the agent is a customer. Narrow inference results in the agent overestimating the price of the good, which hurts the seller as it means they sell to fewer types at any price schedule.

We can then see how the principal's optimal thresholds differ when we move to narrow agents from the rational benchmark. Under narrow inference, if actions have a predictable cost to the agent but benefit to the principal, the principal implements a strategy with a lower type threshold for taking the action on any dimension than in the rational benchmark. The opposite holds in the predictable utility benefit, principal's loss case where the thresholds are higher when the agent is narrow.

Proposition 4. *Assume the IVV assumption holds.*

1. *When for every $i \in N$ we have $w_i > 0$ and $v_i(s) \leq 0$ for all $s \in S$, then on each dimension the objective-maximizing thresholds are weakly lower with narrow agents than under the rational benchmark, so $a_i = 1$ is taken by a larger proportion of types for all $i \in N$.*

2. *When for every $i \in N$ we have $w_i < 0$ and $v_i(s) \geq 0$ for all $s \in S$, then on each dimension the objective-maximizing thresholds are weakly greater with narrow agents than under the rational benchmark, so $a_i = 1$ is taken by a smaller proportion of types for all $i \in N$.*

Proof. In Appendix □

Following the same intuition as for Proposition 3, in the first case narrow inference reduces the marginal cost to the principal of implementing that any given proportion of agents take the action on any dimension. Given the fixed benefits of the actions to the principal, this lower marginal cost means the principal wants a higher proportion of agents to take the action. In the second case, where actions predictably cost the principal but benefit the agent the principal has a lower marginal benefit from a higher proportion of agents taking the action, but a fixed cost. The principal then wants to reduce the proportion of agents taking the actions on all dimensions.

Suppose that whether the agent does narrow inference is something that the principal can influence or design. This could be either through the way they present the mechanism or data about the mechanism. If the default was the agent does narrow inference, the principal could also try to educate the agent on how to do rational inference. Proposition 3 then shows in what cases the principal would gain from de-biasing the agent and in what cases they would not. In settings like the firm-worker story, where the agent makes costly investments that benefit the principal, then the principal wants to salami slice the agent's investment choices into smaller decision problems as much as possible. Conversely, in settings like the buyer-seller story where the agent buys valuable goods that are costly for the principal to produce, the principal wants to merge choices into few decisions. This presents a novel motivation for the seller to engage in bundling the goods they are selling, that aims to mitigate the costs of the buyers agent's bounded rationality to the seller rather using monopoly power to exploit the buyer.[4]

4.4 Effect of Greater Dimensionality

I now consider what happens when the number of dimensions grows large, with symmetry across dimensions. In this setting when actions have a direct benefit for the principal, the principal is able to incentivize types to take the actions at vanishing cost. This is because they only have to pay the transfer to the agent on

[4]See Armstrong (2016) for a review of the literature on nonlinear pricing and bundling.

one dimension in this symmetric narrow agent setting. When actions have a direct cost to the principal, the opposite is true and it becomes too costly for the principal to extract transfers from the agent. In this case, narrow agents overestimate the transfer cost to themselves of taking the action for any given transfer function.

Define a *symmetric dimension space of size* n as follows. For any n, $v_i^{(n)}(s) = \frac{1}{n}v(s)$, and $w_i = \frac{1}{n}w$. Both predictable utility and the principal's direct utility from actions are decreasing as the dimensionality of the action space n grows, but such that the total effect of actions $\sum_{i \in N} v_i^{(n)}(s) = v(s)$, $\sum_{i \in N} w_i^{(n)} = w$ is constant. Let $\hat{s}_i^{(n)}$ be the solution to the principal's problem with narrow agents when there is a symmetric dimension space of size n.

Proposition 5. *Assume the IVV assumption holds. Consider a sequence as $n \to \infty$ of symmetric dimension spaces of size n.*

1. *When $w > 0$, there exists an $\overline{n}$ such that for any $n \geq \overline{n}$, we have that the principal's optimal mechanism with narrow agents implements a strategy such that all types take the action on all dimensions; for all $i \in N$, $\hat{s}_i^n = 0$.*

2. *When $w < 0$, there exists an $\overline{n}$ such that for any $n \geq \overline{n}$, we have that the principal's optimal mechanism with narrow agents implements a strategy such that all types take no action on all dimensions; for all $i \in N$, $\hat{s}_i^n = 1$.*

Proof. In Appendix □

The result follows from the logic discussed in the intuition for the characterization result in Section 4.1. To implement symmetric thresholds, the principal must scale down the transfer utility from taking the action on any dimension by $\frac{1}{n}$ relative to the rational benchmark to maintain the same thresholds as narrow incentive compatible. As n grows large, the contribution of the transfer to the principal's utility then shrinks and the transfer transfers to or from the agent become smaller. Eventually for some n, the direct predictable utility to the principal dominates their objective. When $w > 0$ this means they want all types of agent to take the beneficial action while when $w < 0$ they want no types to take the action.

Proposition 5 provides a theory of organization design. Let the principal be the CEO or top management of a multi-divisional firm, and the agent be the firms divisions. The transfer represents the reinvestment in the firm the CEO allows from overall firm profits, where profits $\sum_{i\in N} w_i$ are the CEO's predictable utility from actions taken by the divisions. The divisions have common preferences and obtain predictable utility from their actions and utility from greater reinvestment in the firm. Each division makes narrow inference about how their own productive actions affect total reinvestment in the firm by the CEO. If divisions take actions that are predictably costly to themselves but benefit the firms overall profitability, then under narrow inference the CEO gains from splitting divisions. Conversely, if the divisions take actions that predictably benefit themselves but reduce profitability, then the CEO would want to merge divisions under narrow inference. Under rational inference whether the divisions are merged or split would not affect the CEO's welfare.

4.5 Symmetric Predictable Utility

If the predictable utilities of each action to the agent are the same across dimensions, we have that the threshold strategy implemented by the principal's optimal mechanism has an identical threshold in every dimension. This means the agent perfectly correlates their actions across dimension; above some threshold type the agent chooses action $a_i = 1$ for all $i \in N$ and below this threshold the agent chooses $a_i = 0$. This contrasts with the principal's optimal mechanism in the rational benchmark, where the principal generally implements thresholds that differ across dimensions even when predictable utilities are symmetric.

Proposition 6. *Assume the IVV assumption holds. If predictable utility of the actions is the same across all dimensions; $v_i(s) = v_j(s) = v(s)$ for any $i, j \in N$, and either*

1. *For every $i \in N$, we have $w_i > 0$ and $v(s) \leq 0$ for all $s \in S$.*

2. *For every $i \in N$, we have $w_i < 0$ and $v(s) \geq 0$ for all $s \in S$.*

Then the principal's optimal mechanism with narrow agents implements a threshold strategy such that for any two dimensions $i, j \in N$, $\hat{s}_i = \hat{s}_j$.

Proof. In Appendix ◻

The result holds because with symmetric predictable utilities, then in both cases the narrow participation constraints have to bind on all dimensions. The statistical correctness constraint then pins down that the thresholds have to be equal across dimensions.

Intuitively, if two thresholds differ then under symmetric predictable utilities the principal can either equalise the higher threshold to the lower one when they benefit from the actions ($w_i > 0$), or raise the the lower threshold to the higher level if the actions are costly to the principal ($w_i < 0$). Under the assumptions on the agent's predictable utilities for both cases, this can be done in a way that improves the principal's direct utility without changing the expected transfer utility costs or benefits to the principal.

4.6 Illustrative Example

The following example illustrates the rational benchmark and the narrow agent results.

Example 2. We return to the firm-worker story. A worker faces two human capital investment decisions $N = \{1, 2\}$. They can choose to obtain technical skills $a_1 = 1$ or not $a_1 = 0$ and/or to obtain managerial skills $a_2 = 1$ or not $a_2 = 0$. The disutility of the worker from obtaining the dimension i skills is given by $v_i(s) = -r_i(1 - s)$. This varies with a uniformly distributed type; $s \sim U[0, 1]$, $P(s) = s$. These skills have equal benefit for the firm that employs the worker; $w_1 = w_2 = w > 0$. Assume that it is more costly for the agent to acquire management skills than to acquire technical skills, $r_2 \geq r_1 > 0$.

For these parameters, we have that for each $i \in N$

$$\phi_i(s) = v_i(s) - \frac{1 - P(s)}{p(s)} v_i'(s) = -2r_i(1 - s) \tag{24}$$

We have increasing virtual values and the strategy implemented by the principal's optimal mechanism has a threshold form $g_i(1|s) = \mathbb{1}\{s \geq \hat{s}_i^*\}$ in each dimension, with thresholds

$$\hat{s}_i^* = (1 - \frac{w}{2r_i})\mathbb{1}\{w \leq 2r_i\} \tag{25}$$

This illustrates some simple corollaries of the principal's optimal mechanism in the rational benchmark given by Proposition 1. The threshold in any dimension only depends on the predictable utilities for that dimension, and is decreasing in the size of the predictable benefit to the principal w and increasing in the predictable cost to the agent r_i. These thresholds give the principal objective value

$$W^{rat} = \sum_{i \in N} [\frac{w^2}{4r_i}\mathbb{1}\{w \leq 2r_i\} + (w - r_i)\mathbb{1}\{w > 2r_i\}] \tag{26}$$

With an agent who does narrow inference, given β, the thresholds that solve the problem $\max_{\hat{s} \in [0,1]^n} \overline{W}(\hat{s}, \beta)$ are

$$\hat{s}_i(\beta_i) = (1 - \frac{w}{2r_i\beta_i})\mathbb{1}\{w \leq 2\beta_i r_i\} \tag{27}$$

for $i \in N$. These are always weakly greater than the rational thresholds (25), which demonstrates Proposition 4.

From Theorem 1, we can solve for the principal's optimal thresholds and the value of the principal's objective. There are four separate cases

1. If $w \geq 2r_2$ then $\hat{s}_1 = \hat{s}_2 = 0$ is optimal for the principal, and the principal obtains welfare

$$W_1 = -r_2 + 2w$$

2. If $2r_2 > w \geq 2\sqrt{r_1 \cdot r_2}$ then $\hat{s}_1 = 0, \hat{s}_2 = 1 - \frac{w}{2r_2}$ is optimal for the principal,

and the principal obtains welfare

$$W_2 = \frac{w^2}{4r_2} + w$$

3. If $2\sqrt{r_1 r_2} > w \geq \frac{2\sqrt{r_1 \cdot r_2}}{1+\sqrt{\frac{r_2}{r_1}}}$ then $\hat{s}_1 = 0$, $\hat{s}_2 = 1 - \sqrt{\frac{r_1}{r_2}}$ is optimal for the principal, and the principal obtains welfare

$$W_3 = -r_1 + \sqrt{\frac{r_1}{r_2}} w + w$$

4. If $\frac{2\sqrt{r_1 \cdot r_2}}{1+\sqrt{\frac{r_2}{r_1}}} > w \geq 0$ then $\hat{s}_1 = 1 - \frac{w}{2r_1}(1 + \sqrt{\frac{r_1}{r_2}})$, $\hat{s}_2 = 1 - \frac{w}{2r_2}(1 + \sqrt{\frac{r_2}{r_1}})$ is optimal for the principal, and the principal obtains welfare

$$W_4 = \frac{w^2}{4}(\frac{1}{\sqrt{r_1}} + \frac{1}{\sqrt{r_2}})^2$$

We can compare the principal's objective value under the rational benchmark W^{rat} with the objective value for narrow agents in any of the four regimes. We can verify that for the parameter values under which each case is optimal, the principal's welfare is higher under narrow inference than in the rational benchmark. This demonstrates Proposition 3.

Note also that the agent's predictable utility in one dimension can affect the optimal threshold in the other dimension. This is particularly clear when looking at the third case, where for a small increase in the cost of acquiring technical skills r_1 actually can lead to an increase in the proportion of types the principal wants to obtain managerial skills. This is in contrast to in the rational benchmark, where under IVV the predictable utility of both the agent and the principal only affect the threshold in their own dimension and where less predictable utility for any action means a higher threshold is optimal for the principal. △

5 Extensions

5.1 Sum-Narrow Participation Constraints

In this section I consider an alternative to narrow dimension-by-dimension participation constraints. The *sum-narrow participation constraint* requires that $\sum_{i\in N} \overline{U}_i(s)$ is greater that the value of the outside option for all types $s \in S$. This is a relaxation of the narrow participation constraint considered earlier. I show how we can modify the analysis of the principal's optimal mechanism for this setting.

The sum-narrow participation constraint holds if for all $s \in S$

$$\sum_{i\in N} \overline{U}_i(s) = \sum_{i\in N} \sum_{a_i\in A_i} g_i(a_i|s)[v_i(s)a_i + \overline{t}_i(a_i)] \geq 0 \tag{28}$$

This constraint reflects that the agent makes a joint decision across dimensions on whether to participate or not. The agent has the option to take the zero action on all dimensions and reject any transfer given by the mechanism. Under the sum-narrow participation constraint, the agent understands that the transfer function may be interactive across dimensions and they cannot just reject the transfer on each dimension individually. This contrasts with their beliefs formed from narrow inference for different actions within the mechanism, which could be correct only if the transfer function was additive.

I modify the proof of Theorem 1 to obtain the following result.

Theorem 2. *Assume IVV holds. The principal maximizes their objective over all NIC mechanisms that satisfy the sum-narrow participation constraint if and only if they choose a transfer function that implements a threshold strategy that solves*

$$\max_{\hat{s}\in[0,1]^n} \overline{W}(\hat{s}; \frac{1}{n}) = \max_{\hat{s}\in[0,1]^n} \{\sum_{i\in N} \int_{\hat{s}_i}^{1} (\frac{1}{n}\phi_i(s) + w_i)\, p(s)ds\} \tag{29}$$

Proof. In Appendix □

The only subsequent result that doesn't hold under sum-narrow participation constraints is Proposition 6. Propositions 3, 4 and 5 continue to hold, and their proofs require no modification.

5.2 Dropping the IVV assumption

I now explore what happens to the principal's optimal mechanism without the IVV assumption. I first present an example where IVV does not hold, and show the principal's optimal mechanism under the rational benchmark is no longer separable and no longer implements a deterministic threshold strategy. Thus, since only deterministic threshold strategies are NIC, we cannot make the same neat comparisons between the principal's optimal mechanism under narrow and rational inference that we made in Propositions 3 and 4. Without IVV, the fact that there are strategies that are IC but not NIC becomes relevant.

However, despite the reduction in the set of strategies that are implementable we can prove that an analogous result to Proposition 3 still holds. The principal is still always better off when agents make narrow compared to rational inference if $v_i(s) \leq 0$, $w_i > 0$ for all $i \in N$, $s \in S$ and the principal is restricted to implement a *deterministic interval strategy*. In a deterministic interval strategy, for some $k \in \mathbb{N}$ there is a partition of the type space $z_0 = 0 < z_1 < ... < z_{k-1} < z_k = 1$ where for any $s \in [z_{l-1}, z_l)$ for $l \in \{1, ..., k-1\}$ (and $[z_{k-1}, 1]$ for $l = k$) we have $g(a|s) = 1$ for some $a \in A$.

Finally, I show how the characterization result Theorem 1 can be modified to deal with the absence of IVV.

5.2.1 Example without IVV

Example 3. There are two dimensions $N = \{1, 2\}$ and the type distribution is uniform $P(s) = s$. Let the predictable utility take the following form on either dimension, for some d_i, b_i and $r_i \in (0,1)$

$$v_i(s) = \begin{cases} (d_i - b_i)(1-r_i) - \frac{d_i}{2}(1-s) - \frac{d_i - b_i}{2}\frac{(1-r_i)^2}{1-s} \text{ if } s \in [0, r_i) \\ -\frac{b_i}{2}(1-s) \text{ if } s \in [r_i, 1] \end{cases} \tag{30}$$

this has a continuous derivative equal to

$$v_i'(s) = \begin{cases} \frac{d_i}{2} - \frac{d_i - b_i}{2}(\frac{1-r_i}{1-s})^2 \text{ if } s \in [0, r_i) \\ \frac{b_i}{2} \text{ if } s \in [r_i, 1] \end{cases} \tag{31}$$

Assume that $d_i(1-(1-r_i)^2) + b_i(1-r_i)^2 > 0$ and $b_i > 0$ so that $v_i(.)$ is strictly increasing. We can write the virtual values associated with these predictable utilities.

$$v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s) = \begin{cases} (d_i - b_i)(1-r_i) - d_i(1-s) \text{ if } s \in [0, r_i) \\ -b_i(1-s) \text{ if } s \in [r_i, 1] \end{cases} \tag{32}$$

These virtual values can be decreasing for some interval of types depending on the parameters. Let $d_1 = b_1 = 0.54$, $r_1 = 0.6$, $d_2 = -0.12$, $b_2 = 1.1$, $r_2 = 0.66$, $w_1 = 0.4266$ and $w_2 = 0.32$. I show that under these parameters the principal's optimal separable mechanism is dominated by a non-separable mechanism implementing a non-threshold strategy.

First calculate the best case mechanism for the principal facing a rational agent if they were restricted to separable mechanisms. For these parameters, the optimal thresholds are interior and solve $\phi_i(\hat{s}_i) + w_i = 0$ for $i \in N$.

$$\hat{s}_1^* = 1 - \frac{w_1}{b_1} = 0.21$$
$$\hat{s}_2^* = 1 - \frac{w_2}{b_2} = \frac{39}{55}$$

The value of the principal's objective under this mechanism is then $W(\hat{s}_1^*, \hat{s}_2^*) \approx 0.215$.

The following mechanism is better for the principal. It implements the deterministic interval strategy g^{int}

$$
\begin{aligned}
&g^{int}(a_1, a_2|s)\\
&= \mathbb{1}\{s \in [0, \hat{s}_1^*)\}(1-a_1)\cdot a_2 + \mathbb{1}\{s \in [\hat{s}_1^*, \hat{s}_2^*)\}a_1\cdot(1-a_2) + \mathbb{1}\{s \in [\hat{s}_2^*, 1]\}a_1\cdot a_2
\end{aligned}
$$

This gives the principal payoff

$$
\begin{aligned}
W(g^{int}) = &\int_0^{\hat{s}_1^*} (\phi_2(s) + w_2)\, p(s) ds + \int_{\hat{s}_1^*}^{\hat{s}_2^*} (\phi_1(s) + w_1)\, p(s) ds\\
&+ \int_{\hat{s}_2^*}^{1} ((\phi_1(s) + w_1) + (\phi_2(s) + w_2))\, p(s) ds\\
&\approx 0.217
\end{aligned}
$$

which is greater that the payoff from the best-case for a separable mechanism. The strategy g^{int} can be made both IC and to satisfy the participation constraint by the following non-separable transfer function.

$$
\begin{aligned}
t(0,0) &= 0\\
t(0,1) &= t(0,0) - v_1(0)\\
t(1,0) &= t(0,1) - v_1(\hat{s}_1^*)\\
t(1,1) &= t(1,0) - v_2(\hat{s}_2^*)
\end{aligned}
$$

$\triangle$

5.2.2 Welfare of the Principal without IVV

I now show that Proposition 3— which distinguishes cases where the principal benefits from facing a narrow agent— also holds without the IVV assumption. In the case where the principal is worse-off under narrow inference, the result is trivial as the principal may now benefit from implementing a larger set of strategies in the rational case compared to the narrow case. The interesting case is the one where the principal gains from narrow inference, where actions have a predictable cost to the agent but a direct benefit to the principal.

The proof for the second case works as follows. For a fixed interval strategy, for each dimension $i \in N$ take the lowest type $\underline{s}_i$ that both takes the action $a_i = 1$ and would produce positive value to the principal if implemented as a threshold; $(v_i(\underline{s}_i) + w_i)(1 - P(\underline{s}_i)) \geq 0$. We then take the dimension i^* at which such a threshold has the greatest cost to the principal in terms of the transfer function required for IC. We can obtain an upper bound on the loss of payoff to the principal if they moved to a threshold strategy where only the action on dimension i^* is taken, and only by types above threshold $\underline{s}_{i^*}$. Under narrow inference, for exactly the same cost the principal has to pay to implement this solo action strategy under the rational benchmark, the principal can implement a threshold strategy where substantial payoff from other dimensions is obtained for free. The gain from this move to narrow inference exceeds the upper bound on the loss from moving to the solo action strategy. This shows the principal's payoff from any interval strategy in the rational benchmark is lower that their payoff from some threshold strategy under narrow inference.

Proposition 7. *Assume the principal is restricted to implementing a deterministic interval strategy*

1. *When for every $i \in N$ we have $v_i(s) \leq 0$ for all $s \in S$, the principal can obtain at least as high an objective value when the agent is narrow compared to the rational benchmark.*

2. *When for every $i \in N$ we have $v_i(s) \geq 0$ for all $s \in S$, the principal obtains at least as high an objective value in the rational benchmark compared to when the agent is narrow.*

Proof. In Appendix □

5.2.3 Minmax characterization without IVV

In this section I show how we can adapt Theorem 1 when IVV does not necessarily hold. This works by applying the ironing procedure of Myerson (1981). Let $P^{-1}(s)$

be the quantile function for the type distribution.[5] For any dimension $i \in N$ and any $x \in [0,1]$, define

$$\Phi_i(x) = \int_x^1 \phi_i(P^{-1}(u))du \tag{33}$$

Let $co(\Phi_i)$ be convex hull of the graph of Φ_i, the upper concave envelope of $\Phi_i(.)$ is then defined as

$$\widehat{\overline{\Phi}}_i(x) = \sup\{z : (z, x) \in co(\Phi_i)\} \tag{34}$$

Then let $\widehat{\overline{\phi}}_i(s) = -\widehat{\overline{\Phi}}'_i(P(s))$ and $\widehat{\overline{\Phi}}_i(x) = \int_x^1 \widehat{\overline{\phi}}_i(P^{-1}(x))$, this is well defined and increasing for all $s \in S$ and $x \in [0,1]$ by definition of $\widehat{\overline{\Phi}}_i(.)$.[6] For any $\beta \in [0,1]^n$ and dimension $i \in N$, define the threshold[7]

$$\hat{s}_i^*(\beta) = \min\{\tilde{s} \in \arg\max_{\hat{s}_i \in S} \int_{\hat{s}_i}^1 (\beta_i \widehat{\overline{\phi}}_i(s) + w_i)\, p(s)ds\}$$

and let $\hat{s}^*(\beta) = (\hat{s}_i^*(\beta))_{i \in N}$. We can now state the result.

Theorem 3. *The principal maximizes maximizes their objective over all NIC mechanisms that satisfy the narrow participation constraints if and only if they choose a transfer function that implements a threshold strategy $\hat{s}^*(\beta^*)$, where*

$$\beta^* \in \arg\min_{\beta \in [0,1]^n : \sum_{i \in N} \beta_i = 1} \sum_{i \in N} \int_{\hat{s}_i^*(\beta)}^1 (\beta_i \widehat{\overline{\phi}}_i(s) + w_i)\, p(s)ds \tag{35}$$

and the value of the principal's objective is given by

$$\widehat{\overline{W}}(\hat{s}^*(\beta^*); \beta^*) = \sum_{i \in N} \int_{\hat{s}_i^*(\beta^*)}^1 (\beta_i^* \widehat{\overline{\phi}}_i(s) + w_i)\, p(s)ds \tag{36}$$

Proof. In Appendix □

[5] This is the inverse of cdf P as P is strictly increasing, and thus P^{-1} is also strictly increasing.

[6] We have that by concavity $\widehat{\overline{\Phi}}'_i(P(.))$ is defined for all but a countable set of points in S, and by right continuity we can extend it to all S.

[7] This is well defined due to continuity of $\int_{\hat{s}_i}^1 (\beta_i \widehat{\overline{\phi}}_i(s) + w_i)\, p(s)ds$.

5.3 Connection to ABEE

It is possible to express behaviour under narrow inference as an Analogy Based Expectation Equilibrium (ABEE) (Jehiel, 2005). Under ABEE, each player in a game has an 'analogy partition' of the set of histories where other players move. For any cell in the partition, a player believes that the strategy of the other players is the average of the true strategies for histories in that cell.

Take a game with $n+2$ players; consisting of the principal, n different 'selves' of the agent and a player of nature. Each of the n selves corresponds to one of the n actions available to the agent, so that self $i \in \{1, ..., n\}$ controls action a_i. All selves share identical preferences over the actions and transfer. The timing of the game is as follows; first the principal chooses a transfer function t. Then the common type of the agent's selves is drawn. After learning this common type then, moving in any order, each of the n selves choose either an action from the set they control or to not participate in the mechanism. Finally, the player of nature implements the transfer function chosen by the principal.

Although each of the agent's selves have common preferences, they differ according to their analogy partitions. Each self partitions the history at which the player of nature moves, with each cell in the partition corresponding to a different action chosen by the self. Thus their beliefs about the expected transfer from each action is the average transfer obtained among all types of agents choosing that action. This coincides with the beliefs under narrow inference. Behaviour under narrow inference then coincides with an ABEE of this multi-selves game.

6 Conclusion

This paper takes a step towards understanding how errors in causal inference might affect how we should design economic incentives. I consider a model boundedly rational belief formation I call narrow inference. I then explore the consequences of this model for economic design. In Theorem 1, I show how we can solve for a principal's optimal mechanism when facing an agent who makes narrow inference, and how this mechanism contrasts with the principal's optimal mechanism when they face an agent who is fully rational. I demonstrate how differences in the underlying environment affect both whether the principal benefits or not from causal inferential errors and the shape of the principal's favoured mechanism. I also demonstrate the robustness of these conclusions to some variations of the underlying model.

One can imagine many additional variations and extensions of the model of bounded rationality explored in this paper. In particular, it seems interesting to consider how the principal could shape the extent of agent's departure from rational beliefs. For example, the principal could provide data on how additional variables correlate with the transfers or otherwise frame the mechanism in a way that influences inference by the agent. The analysis of this paper suggests that in some cases this could be as important a margin of design as the size of material incentives.

A Appendices

A.1 Rational Agents

We analyze the principal's optimal mechanism in the rational benchmark. The next result provides a standard characterization of all IC strategies and expected utilities. We say that expected utilities $\{U(s)\}_{s\in S}$ can be *achieved* given strategy g if there exists a transfer function t such that for every type $s \in S$, $U(s)$ is the expected utility.

Lemma A.1. *A strategy g and expected utilities $\{U(s)\}_{s\in S}$ that can be be achieved given g are IC only if*

1. *Weak monotonicity condition: For any $s, s' \in S$*

$$\sum_{i\in N}(v_i(s) - v_i(s')) \sum_{a_i\in A_i} g_i(a_i|s)a_i \geq \sum_{i\in N}(v_i(s) - v_i(s')) \sum_{a_i\in A_i} g_i(a_i|s')a_i \tag{37}$$

2. *The expected utility $U(s)$ is increasing in $s \in S$.*

3. *The following envelope condition holds for any two types $s, s' \in S$*

$$U(s) = U(s') + \sum_{i\in N} \int_{s'}^{s} v_i'(z) \sum_{a_i\in Y} a_i g(a_i|z)dz \tag{38}$$

Proof. For any types $s, s' \in S$, the rational incentive constraints (5) require that

$$U(s) \geq U(s') + \sum_{i\in N}(v_i(s) - v_i(s')) \sum_{a_i\in A_i} g_i(a_i|s')a_i$$

Together with the IC from interchanging s, s' in the above, we get that the weak monotonicity condition must hold. This then implies the second condition. Using the rewritten ICs, the envelope condition holds from the Lipschitz continuity arguments in Theorems 1 and 2 of Milgrom and Segal (2002). □

As noted, a strategy where the expected utility of actions in individual dimensions is non-monotonic in type can be IC as long as the weak monotonicity condition is satisfied. This is in contrast to the narrow agent model where the action has to be monotonic in type for each dimension. We shall see that under IVV, this does not matter as the principal's optimal mechanism implements a threshold strategy that is monotone on each dimension anyway.

The example in Section 3.2 of Carroll (2017) shows that in we can have non-separability in an optimal selling mechanism with a co-monotonic type distribution. This is due to different monotonicity condition when we have multiple goods relative to when we have a single good. With a single good, we have that under IC higher types must get the good with higher probability, while with multiple goods we can trade-off probabilities across goods without violating IC. In the example of Section 3, we show that this also applies in our model when IVV does not hold.

We use Lemma A.1 to prove the next lemma, showing that any deterministic threshold strategy can be implemented with an additively separable transfer function.

Lemma A.2. *Let g^* be a deterministic threshold strategy and $U(0)$ be the expected utility of the type $s = 0$. The strategy is IC and achieves the expected utility $U(0)$ for type $s = 0$ under transfer function*

$$t(a_1, ..., a_n) = \sum_{i \in N} t^i(a_i) \tag{39}$$

$$t^i(0) = \frac{1}{n} U(0), t^i(1) = -v_i(\hat{s}_i) + \frac{1}{n} U(0) \text{ for all } i \in N \tag{40}$$

Proof. The expected utility of type s behaving according to deterministic threshold strategy g^* when the transfer function is that given in the lemma statement is

$$\begin{aligned} U(s) &= \sum_{i \in N} (v_i(s) + t^i(1) - t_i(0))\mathbb{1}[s \geq \hat{s}_i] + \sum_{i \in N} t^i(0) \\ &= \sum_{i \in N} (v_i(s) - v_i(\hat{s}_i))1[s \geq \hat{s}_i] + U(0) \end{aligned}$$

We can see from this that the threshold strategy is IC, as for any dimension $i \in N$ a type above the threshold $\hat{s}_i$ gets a weakly positive utility from choosing $a_i = 1$

over $a_i = 0$ while a type below the threshold gets a negative utility. The lowest type $s = 0$ gets utility $U(0)$. $\square$

A.1.1 Proof of Proposition 1

Proof. We can rewrite the principal's objective (2) using the envelope formula (38) from Lemma A.1 and the expression for the transfer function in terms of utilities in the direct mechanism.

$$
\begin{aligned}
W(t,g) &= \int_0^1 \sum_{a\in A} [-t(a) + \sum_{i\in N} w_i a_i] g(a|s)\, p(s) ds \\
&= \sum_{i\in N} \int_0^1 \sum_{a_i\in A_i} [a_i v_i(s) + w_i a_i] g_i(a_i|s)\, p(s) ds - \int_0^1 U(s)\, p(s) ds \\
&= \int_0^1 \sum_{a\in A} [a_i v_i(s) + \sum_{i\in N} w_i a_i] g(a|s)\, p(s) ds \\
&- \sum_{i\in N} \int_0^1 [\int_0^s v_i'(z) \sum_{a_i\in A_i} a_i g_i(a_i|z) dz]\, p(s) ds - U(0) \\
&= \int_0^1 \sum_{a\in A} [\sum_{i\in N} a_i v_i(s) + \sum_{i\in N} w_i a_i] g(a|s)\, p(s) ds \\
&- \sum_{i\in N} \int_0^1 [\int_z^1 p(s) ds] v_i'(z) \sum_{a_i\in A_i} a_i g_i(a_i|z) dz - U(0) \\
&= \sum_{i\in N} \int_0^1 \sum_{a_i\in A_i} [a_i v_i(s) + w_i a_i] g_i(a_i|s)\, p(s) ds \\
&- \sum_{i\in N} \int_0^1 [1 - P(z)] v_i'(z) \sum_{a_i\in A_i} a_i g_i(a_i|z) dz - U(0) \\
&= \sum_{i\in N} \int_0^1 \sum_{a_i\in A_i} [\phi_i(s) a_i + w_i a_i] g_i(a_i|s)\, p(s) ds - U(0)
\end{aligned}
$$

Where the first line follows from expressing the transfer function in terms of expected utility and the last two lines follow from a standard switching of the order of integration and rewriting in terms of marginal strategies.

Clearly it is optimal to set the expected utility of the lowest type to zero. We now consider a relaxed version of the Principal's problem where we ignore the weak monotonicity constraints from Lemma A.1 and that expected utilities might not be achieved given g. We show that under IVV, the mechanism that solves this relaxed problem implements a deterministic threshold strategy. Under a threshold strategy we have that for all $i \in N$, $\sum_{a_i \in A_i} a_i g_i(a_i|s)$ is increasing in type s. Thus the weak monotonicity condition of Lemma A.1 is satisfied.

By Lemma A.2 we can find a transfer function that implements the threshold strategy and achieves any given expected utility for the lowest type. Thus the solution to the relaxed problem coincides with the solution to the full problem.

$$\max_{g_i \in \Delta(A_i)^S} \int_0^1 [\sum_{i \in N, a_i \in A_i} (\phi_i(s) + w_i) a_i g_i(a_i|s)] \, p(s) ds \tag{41}$$

This problem can be solved pointwise by strategy $g_i(a_i|s) = 1$ if and only if $a_i \in \arg\max_{\tilde{a}_i \in A_i} (\phi_i(s) + w_i)\tilde{a}_i$. BY IVV, $\phi_i(s) + w_i$ is strictly increasing in $s \in S$. Thus either we have $\phi_i(\hat{s}_i) + w_i = 0$ for some $\hat{s}_i \in [0, 1]$, or either $\phi_i(s) + w_i < 0$ or $\phi_i(s) + w_i > 0$ for all $s \in S$. In the first case the maximizing strategy on dimension $i \in N$ is $g_i(a_i|s) = \mathbb{1}\{s \geq \hat{s}_i\}$, where without loss of generality we set $g_i(a_i|\hat{s}_i) = 1$ since $\hat{s}_i \in S$ has measure zero. In the other cases we can write the maximizing strategy as having a threshold form with thresholds $\hat{s}_i = 0$ and $\hat{s}_i = 1$ respectively. $\square$

A.2 Proofs

Proof of Lemma 1

For any dimension $i \in N$ and any two types $s, s' \in S$ NIC requires

$$\overline{U}_i(s) \geq \overline{U}_i(s') + (v_i(s) - v_i(s')) \sum_{a_i \in A_i} a_i g_i(a_i|s')$$

From this we can obtain that the first two conditions are necessary. The envelope formula on each dimension then holds via the usual Lipschitz continuity arguments as in Milgrom and Segal (2002).

Conversely, combining the envelope formula with the rewritten NIC gives

$$\int_{s'}^{s} v_i'(z) \sum_{a_i \in A_i} a_i g_i(a_i|z) dz \geq (v_i(s) - v_i(s')) \sum_{a_i \in A_i} a_i g_i(a_i|s')$$

which holds under the monotonicity condition.

Proof of Proposition 2

By the envelope formula (20) of Lemma 1, the beliefs inducing any deterministic threshold strategy must satisfy

$$\bar{t}_i(1) = \bar{t}_i(0) - v_i(\hat{s}_i)$$

for each dimension $i \in N$. Thus NIC and the thresholds pin down beliefs, and any threshold strategy can be rendered NIC by some beliefs.

An outline of the proof is as follows. In Lemma A.3 we show that for any two distinct action combinations that occur under a deterministic threshold strategy, one of the action vectors is weakly larger on all dimensions. For each dimension we can partition the set of action combinations, with each cell consisting of all action combinations that share a common action for that dimension. In Lemma A.4 we then show that for every action combination that occurs with positive probability, there is at least one dimension such that every other action combination in the same partition cell for that dimension has a smaller action taken on at least one of the other dimensions.

We can use this fact to recursively construct a transfer function that implements given beliefs, and we can show that this constructed transfer function is additive for all action combinations that occur with positive probability. Finally we extend this transfer function so that is defined on all action combinations, whilst preserving additivity.

Lemma A.3. *Let $\tilde{g}$ be a deterministic threshold strategy. Then for any $s > s'$ and a'', a' such that $a'' \neq a'$, $\tilde{g}(a''|s'') > 0$ and $\tilde{g}(a'|s') > 0$ only if $a''_j \geq a'_j$ for all $j \in N$.*

Proof. Since $\tilde{g}$ is a deterministic threshold distribution, for any two dimensions $i, j \in N$ there is an $\hat{s}_i$ such that $\tilde{g}_i(1|s) = \mathbb{1}\{s \geq \hat{s}_i\}$ and an $\hat{s}_j$ such that $\tilde{g}_j(1|s) = \mathbb{1}\{s \geq \hat{s}_j\}$. Suppose for that we can find $a'' \neq a'$, $\tilde{g}(a''|s'') > 0$ and

$\tilde{g}(a'|s') > 0$ for some $s'', s' \in S$ such that $a''_i = 1 > a'_i = 0$ for some dimension $i \in N$ but $a''_j = 0 < a'_j = 1$ for another dimension $j \in N \setminus \{i\}$. But then $s \geq \hat{s}_i > s''$ and $s'' \geq \hat{s}_j > s$, a contradiction. □

This implies that the any two action combinations occurring with positive probability under a threshold strategy can be ranked. Denote the set of all action combinations that have positive probability under g by $A(g) = \{a \in A : \exists s \in S$ such that $g(a|s) > 0\}$. Denote the projection of $A(g)$ on dimension $i \in N$ by $A_i(g)$. Define the order $\succ$ so that $a'' \succ_i a'$ if and only if $a''_j \geq a'_j$ for all $j \in N$ with strict inequality for at least one such j. By Lemma A.3 this is a strict total order.

Given our NIC deterministic threshold strategy g, we enumerate the set $A(g) = \{1, ..., |A(g)|\}$ so that $k > l$ means that for $a^k, a^l \in A(g)$, $a^k \succ a^l$. Now we can form a partition of $A(g)$ for each dimension $i \in N$. For each action $a_i \in A_i$, define the set $\mathcal{A}_i(a_i) = \{(a_i, \tilde{a}_{-i}) \in A(g)\}$. This is a partition as $\varnothing \notin \mathcal{A}_i(a_i)$ for any $a_i \in A_i(g)$, $\cup_{\tilde{a}_i \in A_i} \mathcal{A}_i(\tilde{a}_i) = A(g)$ and $\mathcal{A}_i(a''_i) \cap \mathcal{A}_i(a'_i) = \varnothing$ for any $a''_i \neq a'_i \in A_i$.

We can then show that any action combination that occurs with positive probability under an NIC deterministic threshold strategy must be maximal in the partition cell according to the order $\succ$ for at least one dimension.

Lemma A.4. *Given an NIC deterministic threshold strategy g, any $a \in A(g)$ is such that for at least one dimension $j \in N$, $a = (a_j, a_{-j}) \succ \tilde{a} = (a_j, \tilde{a}_{-j})$ for all $\tilde{a} \in \mathcal{A}_j(a_j)$.*

Proof. Suppose this does not hold, then for all $i \in N$, we can find a $\tilde{a}(i) \in \mathcal{A}(a_i)$ such that $\tilde{a}(i) \succ a$. The finite set $\{\tilde{a}(1), ..., \tilde{a}(n)\}$ must contain a member that is minimal in the strict total order $\succ$. Denote this element $\tilde{a}(k)$ for some $k \in N$. Then on all dimensions $j \in N$, $\tilde{a}(k)_j \leq a_j = \tilde{a}(j)_j$, as if $\tilde{a}(k)_j > a_j = \tilde{a}(j)_j$ then $\tilde{a}(j) \not\succ \tilde{a}(k)$ which would contradict the minimality of $\tilde{a}(k)$ in $\{a(1), ..., a(n)\}$. However, $\tilde{a}(k)_j \leq a_j$ for all $j \in N$ contradicts that $\tilde{a}(k) \succ a$. □

Now using this, for any action combination $a^l \in A(g) = \{1, ..., |A(g)|\}$ assign a dimension $\pi(l) \in N$ so that $a^l \succ \tilde{a}$ for any $\tilde{a} \in \mathcal{A}_{\pi(l)}(g)$. Then we can recursively define a transfer function t from the beliefs $\bar{t}$ that render g NIC. For any $k \in \{2, ..., |A(g)|\}$

$$t(a^1) = \bar{t}_{\pi(1)}(a^1_{\pi(1)})$$
$$t(a^k) = \frac{\sum_{a \in \mathcal{A}_{\pi(k)}(a^k_{\pi(k)})} g(a)}{g(a^k)} \bar{t}_{\pi(k)}(a^k_{\pi(k)}) - \sum_{a \in \mathcal{A}_{\pi(k)}(a^k_{\pi(k)}) \backslash \{a^k\}} \frac{g(a)}{g(a^k)} t(a)$$

This transfer function is well defined as for any $a^k \in A(g)$ all $a \in A_{\pi(k)}(a^k_{\pi(k)})$, $t(a)$ has been defined at an earlier stage as a^k is maximal in $\succ$ on dimension $\pi(k)$. Since a^1 is minimal in $A(g)$, we have that $a^1 = A_{\pi(1)}(a^1_{\pi(1)})$, so the first equation is in fact a special case of the second. We now show that t is additive for the action combinations $a \in A(g)$ at which it is well defined.

Lemma A.5. *For any $a_{-i} \in A_{-i}$ with $(1, a_{-i}), (0, a_{-i}) \in A(g)$, there exists no $\tilde{a}_{-i} \neq a_{-i}$ such that $(1, \tilde{a}_{-i}) \in A(g)$ and $(0, \tilde{a}_{-i}) \in A(g)$.*

Proof. Suppose for contradiction that there is a $\tilde{a}_{-i} \neq a_{-i}$ such that $(1, \tilde{a}_{-i}) \in A(g)$ and $(0, \tilde{a}_{-i}) \in A(g)$. As we have strict total order $\succ$ on $A(g)$, we have two cases. In the first case $(0, a_{-i}) \succ (0, \tilde{a}_{-i})$. This means that $a_j \geq \tilde{a}_j$ for all $j \in N \backslash \{i\}$ with strict inequality for some such j. Then neither $(0, a_{-i}) \succ (1, \tilde{a}_{-i})$ nor $(1, \tilde{a}_{-i}) \succ (0, a_{-i})$. This is a contradiction as since $(1, \tilde{a}_{-i}) \neq (0, a_{-i})$, Lemma A.3 implies they must be ranked.

Similarly if $(0, \tilde{a}_{-i}) \succ (0, a_{-i})$, then $\tilde{a}_j \geq a_j$ for all $j \in N \setminus \{i\}$ with strict inequality for some such j. Then neither $(1, a_{-i}) \succ (0, \tilde{a}_{-i})$ or $(0, \tilde{a}_{-i}) \succ (1, a_{-i})$. □

Therefore we cannot have $t(1, a_{-i}) - t(0, a_{-i}) \neq t(1, \tilde{a}_{-i}) - t(0, \tilde{a}_{-i})$ and $(1, a_{-i}), (0, a_{-i}), (1, \tilde{a}_{-i}), (0, \tilde{a}_{-i}) \in A(g)$. This means the transfer function is additive for all $a \in A(g)$.

We can additively extend the transfer function t defined above to all $a \in A$. Denote this extended transfer function by t'. For any $\tilde{a} \in A$, such that $a^1_j \geq \tilde{a}_j$ for all $j \in N$, define $t'(\tilde{a}) = t(a^1)$. For each dimension $i \in N$, we will define $t^i(a_i)$ for each $a_i \in A_i$ so that $t'(a) = \sum_{i \in N} t^i(a_i)$. First set $t^i(a^1_i) = \frac{1}{n} t(a^1)$

for all $i \in N$, and $t^i(1) = t^i(0)$ if $a_i^1 = 1$. For dimensions which are such that $(1, a_{-i}) \notin A(g)$ for every $a_{-i} \in A_{-i}$, set $t^i(1) = t^i(0)$.

Now move through the elements $a^l \in \{2, ..., |A(g)|\} \subset A(g)$ in the $\succ$ order. If a^l differs in one dimension j from a^{l-1}, then by Lemma A.5 there is a unique $a_{-j} \in A_{-j}$ such that $(1, a_{-j}), (0, a_{-j}) \in A(g)$, and we can write $t^j(1) - t^j(0) = t(1, a_{-j}) - t(0, a_{-j}) = t(a^l) - t(a^{l-1})$. If a^l differs from a^{l-1} on multiple dimensions (denoted by the set N^l), choose an arbitrary $j \in N^l$ and set $t^j(1) - t^j(0) = t(a^l) - t(a^{l-1})$ and set $t^k(1) = t^k(0)$ for all other $k \in N^l \setminus \{j\}$. This process results in a transfer function that is additive and such that $t'(a) = \sum_{i \in N} t^i(a_i) = t(a)$ for every $a \in A(g)$.

For the final part, any $a \in A \setminus A(g)$ is such that $g(a) = 0$ and thus $t(a)$ does not affect on-path beliefs $\bar{t}(a_i) = \sum_{a_{-i} \in A_{-i}} \frac{g(a_i, a_{-i})}{g_i(a_i)} t(a)$. For all $a \in A(g)$ and t' that implements the beliefs $\bar{t}$ must match our recursive construction t. To see this take $a^1 \in A(g)$, at this action combination $t'(a^1) = t(a^1)$ is pinned down by the beliefs only. This is because a^1 is the minimal action in $A(g)$ according to $\succ$ but is also maximal in $\succ$ on one of the dimension partitions, so on this dimension $t'(a^1)$ and $t(a^1)$ must both be equal to the belief. Then any expression that implements $\bar{t}$ for $t'(a^2)$ is also pinned down in terms of beliefs according to the recursive formula given for $t(a_2)$. Continuing up the order $\succ$, we have $t'(a^l) = t(a^l)$ for every $a^l \in A(g)$.

Proof of Theorem 1

We break the proof into 4 steps.

Step 1: We write the principal's problem in a virtual value form. From the statistical correctness constraint for any dimension $i \in N$ we can write

$$\int_0^1 g(a|s)t(a) = \sum_{a_i \in A_i} g_i(a_i)\bar{t}_i(a_i) = \int_0^1 \sum_{a_i \in A_i} \bar{t}_i(a_i) g_i(a_i|s)\, p(s) ds$$

We then have for any $i, j \in N$ that

$$\int_0^1 \sum_{a_i \in A_i} \overline{t}_i(a_i) g_i(a_i|s)\, p(s) ds = \int_0^1 \sum_{a_j \in A_j} \overline{t}_j(a_j) g_j(a_j|s)\, p(s) ds$$

$$\Leftrightarrow \int_0^1 \overline{U}_i(s)\, p(s) ds - \int_0^1 v_i(s) \sum_{a_i \in A_i} a_i g_i(a_i|s)\, p(s) ds$$

$$= \int_0^1 \overline{U}_j(s)\, p(s) ds - \int_0^1 v_j(s) \sum_{a_j \in A_j} a_j g_j(a_j|s)\, p(s) ds$$

We can then write the principals objective in terms of the perceived utility in one of the dimensions, and then use the envelope formula to write in terms of the utility of the lowest type and the marginal strategy.

$$W(t, g)$$

$$= \int_0^1 \sum_{a \in A} [-t(a) + \sum_{i \in N} w_i a_i] g(a|s)\, p(s) ds$$

$$= - \int_0^1 \overline{U}_i(s)\, p(s) ds + \int_0^1 v_i(s) \sum_{a_i \in A_i} a_i g_i(a_i|s)\, p(s) ds$$

$$+ \sum_{j \in N} \sum_{a_j \in A_j} w_j \int_0^1 a_j g_j(a_j|s)\, p(s) ds$$

$$= -\overline{U}_i(0) - \int_0^1 (\int_0^s v_i(z) \sum_{a_i \in A_i} a_i g_i(a_i|z) dz)\, p(s) ds$$

$$+ \int_0^1 v_i(s) \sum_{a_i \in A_i} a_i g_i(a_i|s)\, p(s) ds + \sum_{j \in N} \sum_{a_j \in A_j} w_j \int_0^1 a_j g_j(a_j|s)\, p(s) ds$$

$$= -\overline{U}_i(0) + \sum_{a_i \in A_i} \int_0^1 (v_i(s) - \frac{1 - P(s)}{p(s)} v_i'(s)) g_i(a_i|s)\, p(s) ds$$

$$+ \sum_{j \in N} \sum_{a_j \in A_j} w_j \int_0^1 a_j g_j(a_j|s)\, p(s) ds$$

Where we use a switch of the order of integration in the final equation. We can also use the envelope formula to write the statistical correctness constraints for any $i, j \in N$ as

$$
\begin{aligned}
&-\overline{U}_i(0) + \sum_{a_i \in A_i} \int_0^1 (v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s)) g_i(a_i|s)\, p(s) ds \\
&= -\overline{U}_j(0) + \sum_{a_j \in A_j} \int_0^1 (v_j(s) - \frac{1-P(s)}{p(s)} v_j'(s)) g_j(a_j|s)\, p(s) ds
\end{aligned}
$$

Step 2: We solve problem that is an upper bound to the principal's full problem and show that the solution to this upper bound implements a deterministic threshold strategy.

The principal wants to maximize $W(t, g)$ and must implement beliefs that satisfy the statistical correctness constraints. The Lagrangian of the problem of maximizing this objective given this constraint can be written as follows, remembering that $\phi_i(s) = v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s)$ and denoting Lagrange multipliers by $\lambda_j \in \mathbb{R}$ for the jth of the $n-1$ statistical correctness constraints

$$
\begin{aligned}
&\overline{W}(g, \overline{U}(0), \lambda) \\
&= -\overline{U}_i(0) + \sum_{a_i \in A_i} \int_0^1 \phi_i(s) a_i g_i(a_i|s)\, p(s) ds \\
&+ \sum_{i \in N} \sum_{a_i \in A_i} w_i \cdot a_i \int_0^1 g_i(a_i|s)\, p(s) ds \\
&+ \sum_{j \in N \setminus \{i\}} \lambda_j [-\overline{U}_j(0) + \sum_{a_j \in A_j} \int_0^1 \phi_j(s) a_j g_j(a_j|s)\, p(s) ds \\
&+ \overline{U}_i(0) - \sum_{a_i \in A_i} \int_0^1 \phi_i(s) a_i g_i(a_i|s)\, p(s) ds] \\
&= (1 - \sum_{j \in N \setminus \{i\}} \lambda_j)[-\overline{U}_i(0) + \sum_{a_i \in A_i} \int_0^1 \phi_i(s) a_i g_i(a_i|s)\, p(s) ds] \\
&+ \sum_{j \in N \setminus \{i\}} \lambda_j [-\overline{U}_j(0) + \sum_{a_j \in A_j} \int_0^1 \phi_j(s) a_j g_j(a_j|s)\, p(s) ds \\
&+ \sum_{i \in N} \sum_{a_i \in A_i} w_i \cdot a_i \int_0^1 g_i(a_i|s)\, p(s) ds
\end{aligned}
$$

Note that we can write the $n-1$ lagrange multipliers as $\beta = (\beta_1, ..., \beta_n) \in \mathbb{R}^n$ with $\sum_{j \in N} \beta_j = 1$ by setting $\beta_j = \lambda_j$ for $j \in N \setminus \{i\}$ and $\beta_i = 1 - \sum_{j \in N \setminus \{i\}} \lambda_j$.

We can use this to write a relaxed version of the principal's problem

$$\sup_{g\in\Delta(A)^S,\overline{U}(0)\in\mathbb{R}^n} \min_{\beta\in\mathbb{R}^n,\sum_{j\in N}\beta_j=1} \overline{W}(g,\overline{U}(0),\beta)$$

$$\text{subject to } \overline{U}_j(0)\geq 0 \text{ for } j\in N$$

The problem is relaxed as it does not satisfy the following constraints that must hold in the full problem. These are that the strategy must satisfy the monotonicity requirement implied by NIC in Lemma 1. In addition since t only depend on actions, for a given strategy g there might not exist a t inducing beliefs $\overline{t}$ such that any arbitrary expected utilities $\overline{U}_i(s)=\sum_{a_i\in A_i} g_i(a_i|s)[a_i v_i(s)+\overline{t}_i(a_i)]$ can be achieved.

We will show that the solution to the relaxed problem implements a strategy that has a deterministic threshold form. Such a strategy satisfies the monotonicity requirement and will be able to induce the expected utilities and beliefs in the solution. Thus both type of additional constraints hold in the solution to the relaxed problem that ignores them.

Restricting the domain of β so that $\beta_j\in[0,1]$ for all $j\in N$ in the minimization problem gives us the following upper bound.

$$\begin{aligned}
&\min_{\tilde{\beta}\in[0,1]^n:\sum_{i\in N}\tilde{\beta}_i=1} \sup_{g\in\Delta(A)^S,\overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \overline{W}(g,\overline{U}(0),\beta)\\
\geq &\sup_{g\in\Delta(A)^S,\overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \min_{\tilde{\beta}\in[0,1]^n:\sum_{i\in N}\tilde{\beta}_i=1} \overline{W}(g,\overline{U}(0),\beta)\\
\geq &\sup_{g\in\Delta(A)^S,\overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \min_{\tilde{\beta}\in\mathbb{R}^n:\sum_{i\in N}\tilde{\beta}_i=1} \overline{W}(g,\overline{U}(0),\beta)
\end{aligned}$$

For fixed $\beta\in[0,1]^n$ with $\sum_{j\in N}\beta_j=1$, we have

$$\overline{W}(g,\overline{U}(0),\beta)=\sum_{j\in N}[-\beta_j\overline{U}_j(0)+\sum_{a_j\in A_j}\int_0^1(\phi_j(s)+w_j)a_j g_j(a_j|s)p(s)ds]$$

The strategy $g\in\Delta(A)^S$ that maximizes this expression must satisfy $g_i(a_i|s)=1$ if and only if $a_i\in\arg\max_{\tilde{a}_i\in A_i}(\beta_i\phi_i(s)+w_i)\tilde{a}_i$ for every $i\in N$. As in Proposition

1, this strategy takes a threshold form WLOG and can be induced by the following beliefs that give the correct expected utilities. Given a dimension $i \in N$, and threshold $\hat{s}_i \in [0,1]$ the following beliefs implement the truthful reporting for the threshold strategy g.

$$\overline{t}_i(0) = -\overline{U}_i(0)$$
$$\overline{t}_i(1) = \overline{t}_i(0) - v_i(\hat{s}_i)$$

By Proposition 2, we can construct a transfer function t that implements these beliefs given the threshold strategy.

Step 3: We show that the objective function in our upper bound problem satisfies the conditions of the minimax theorem. This allows us to interchange the min and max operator and means we have a saddle point solution.

Denote the vector of thresholds by $\hat{s} = (\hat{s}_i)_{i\in N} \in [0,1]^n$. We can now write the objective in terms of the thresholds.

$$\overline{W}(\hat{s}, \overline{U}(0), \beta) = \sum_{j \in N} [-\beta_j \overline{U}_j(0) + \int_{\hat{s}_i}^{1} (\phi_i(s) + w_i)\, p(s) ds]$$

Define the quantile function $P^{-1}(s)$. Since $P(s)$ is strictly increasing, this is just the inverse and is also strictly increasing. For any vector $x \in [0,1]^n$, we can write $P^{-1}(x_i) = \hat{s}_i$. Letting $P^{-1}(x) = (P^{-1}(x_i))_{i\in N}$, we use this to rewrite the objective.

$$\overline{W}(P^{-1}(x), \overline{U}(0), \beta) = \sum_{j \in N} [-\beta_j \overline{U}_j(0) + \int_{x_j}^{1} (\phi_j(P^{-1}(u)) + w_j) du]$$

Taking derivatives of $\int_{x_j}^{1} (\phi_j(P^{-1}(u)) + w_j) du$ with respect to the threshold x_j gives

$$-(\phi_j(P^{-1}(u)) + w_j)$$

By the IVV assumption, this is decreasing and thus $\int_{x_j}^{1} (\phi_j(P^{-1}(u)) + w_j) du$

is concave in $x \in [0,1]^n$. We have that $-\beta_j \overline{U}_j(0)$ is also concave in $x \in [0,1]^n$. The sum of concave functions is also concave, thus for fixed β, $\overline{W}(P^{-1}(x), \overline{U}(0), \beta)$ is concave in $\overline{U}(0)$ and the quantiles x. Since $\overline{W}(P^{-1}(x), \overline{U}(0), \beta)$ is convex in β for fixed $\overline{U}(0), x$ we can apply the minimax theorem (Sion, 1958) and obtain that

$$
\begin{aligned}
&\min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \sup_{x \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \overline{W}(P^{-1}(x), \overline{U}(0), \beta) \\
= &\sup_{x \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \overline{W}(P^{-1}(x), \overline{U}(0), \beta^*) \\
= &\sup_{x \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \overline{W}(P^{-1}(x), \overline{U}(0), \beta)
\end{aligned}
$$

where β^* is the minimizer. Using $\hat{s}_i = P^{-1}(x_i)$, we then have

$$
\begin{aligned}
&\min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \overline{W}(\hat{s}, \overline{U}(0), \beta) \\
= &\sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \overline{W}(\hat{s}, \overline{U}(0), \beta^*) \\
= &\sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \overline{W}(\hat{s}, \overline{U}(0), \beta)
\end{aligned}
$$

Step 4: We can show we can attain the value of the upper bound in the full problem by the following argument. Take the minimizing β^*. If $\beta_i^* > 0$ we must have that when comparing the dimension i and any other dimension j that

$$
\begin{aligned}
&-\overline{U}_i(0) + \int_{\hat{s}_i}^1 (v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s))\, p(s) ds \\
\leq &-\overline{U}_j(0) + \int_{\hat{s}_j}^1 (v_j(s) - \frac{1-P(s)}{p(s)} v_j'(s))\, p(s) ds
\end{aligned}
$$

So that the dimension i term is minimal. At the solution the dimension i participation constraint binds $\overline{U}_i(0) = 0$ as otherwise the principal could increase the value of the objective by reducing $\overline{U}_i(0)$. If $\beta_j^* > 0$, then we have that this inequality must hold with equality with $\overline{U}_j(0) = 0$, which means the statistical correctness and participation constraint hold. If $\beta_j^* = 0$, then we can increase the value of $\overline{U}_j(0)$ without affecting the principal's objective value. From the above inequality, this can be done so that the statistical correctness constraint holds without violating the participation constraint. Thus the solution for the upper bound can be be attained by a NIC strategy and t that satisfy the statistical correctness and participation constraints, and therefore solves the full problem. As $\beta_i^* > 0$, we have that $\overline{U}_i(0) = 0$ so the equality can be achieved with bounded $\overline{U}(0)$. Therefore we can replace the supremum with a maximum in the saddle point problem.

Proof of Proposition 4

For every $i \in N$, let $\hat{s}_i^{rational}$ be the threshold such that $g_i(1|s) = \mathbb{1}\{s_i \geq \hat{s}_i^{rational}\}$ is the strategy that solves the principal's problem in the rational benchmark. Let $\hat{s}_i^{narrow}$ be the solution to the principal's problem a narrow agent, as given by the solution to the minimax problem in Theorem 1. Let β^* be the saddle point distribution over dimensions from that problem.

When $v_i(1) \leq 0$ we have $\phi_i(s) = v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s) \leq 0$ for all $s \in [0, 1]$, then $\phi_i(s) + w_i \leq \beta_i^* \phi_i(s) + w_i$. Suppose that for some dimension $j \in N$ we have $\hat{s}_j^{narrow} > \hat{s}_j^{rational}$. Optimality of $\hat{s}_j^{rational}$ as a threshold implies $\phi_j(s) + w_j > 0$ for all $s \in (\hat{s}_j^{rational}, 1]$. However, then $\hat{s}_j^{narrow} > \hat{s}_j^{rational}$ cannot be optimal for the principal as $0 < \phi_j(s) + w_j \leq \beta_j^* \phi_j(s) + w_j$ for $s \in (\hat{s}_j^{rational}, \hat{s}_j^{narrow}]$, a contradiction.

For the second case we again assume that $\beta_i^* > 0$, as otherwise $\hat{s}_i^{narrow} = 1$ in which case the result holds. Then $\phi_i(s) = v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s) \geq 0$ for any $s \in [\hat{s}_i^{narrow}, 1]$. Otherwise there is a $\tilde{s}_i \in (\hat{s}_i^{narrow}, 1]$ such that by IVV $\beta_i^* \phi_i(s) + w_i < 0$ for all $s \in [\hat{s}_i^{narrow}, \tilde{s}_i]$, and the principal could then switch to implementing $g_i(0|s) = 1$ for all $s \in [\hat{s}_i^{narrow}, \tilde{s}_i]$ and obtain a higher payoff.

Now for contradiction assume $\hat{s}_i^{rational} > \hat{s}_i^{narrow}$. We have that $\phi_i(s) \geq 0$ and thus $\phi_i(s) + w_i \geq \beta_i^* \phi_i(s) + w_i$ for any $s \in [\hat{s}_i^{narrow}, \hat{s}_i^{rational}]$. Then optimality of $\hat{s}_i^{rational}$ is contradicted by the fact that optimality of $\hat{s}_i^{narrow}$ requires $0 < \beta_i^* \phi_i(s) + w_i \leq \phi_i(s) + w_i$ for all $s \in (\hat{s}_i^{narrow}, \hat{s}_i^{rational}]$, where the first strict inequality follows from IVV.

Proof of Proposition 3

For any fixed threshold strategy and fixed β the difference in the principal's objective can be written as

$$
\begin{aligned}
& W(\hat{s}) - \overline{W}(\hat{s};\beta) \\
&= \sum_{i\in N} \int_{\hat{s}_i}^{1} [(\phi_i(s) + w_i) - (\beta_i \phi_i(s) + w_i)]\, p(s)\, ds \\
&= \sum_{i\in N} (1-\beta_i) \int_{\hat{s}_i}^{1} (v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s))\, p(s)\, ds \\
&= \sum_{i\in N} (1-\beta_i) v_i(\hat{s}_i)(1-P(\hat{s}_i))
\end{aligned}
$$

where the last line follows from the fact that $\int_{\hat{s}_i}^{1} (v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s))\, p(s)\, ds = v_i(\hat{s}_i)(1-P(\hat{s}_i))$. Then as $\hat{s}_i \in [0,1]$ for the first case where $v_i(s) \leq 0$ clearly we have $W(\hat{s}) \leq \overline{W}(\hat{s};\beta)$ and for the second case where $0 \leq v_i(s)$ we have $W(\hat{s}) \geq \overline{W}(\hat{s};\beta)$.

Proof of Proposition 6

Let $(\hat{s}^*, \beta^*)$ be the saddle point solution to the characterization problem (16). If for any pair of dimensions $i, j \in N$, $\beta_i^*, \beta_j^* \in (0,1)$ then

$$
\int_{\hat{s}_i}^{1} (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s)\, ds = \int_{\hat{s}_j}^{1} (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s)\, ds
$$

as otherwise we would have $\beta_k^* = 0$ for one of the dimensions $k = \{i, j\}$ as putting any weight on that dimension would not be minimizing.

The remaining case is when $\beta_i^* = 0$ for some $i \in N$. Then it is optimal for the principal to implement threshold $\hat{s}_i^* = 0$ when $w_i > 0$ and $\hat{s}_i^* = 1$ when $w_i < 0$. But then in both cases either all dimensions either choose the same marginal strategy as on i, in which case the result holds, or the solution to the minimization problem would be to have $\beta_i^* > 0$, a contradiction. To see this, in the first case for any $j \in N$ unless $\hat{s}_j^* = 0$ we have.

$$\begin{aligned}
&\int_0^1 (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s) ds \\
&< \int_{\hat{s}_j^*}^1 (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s) ds = v(\hat{s}_j^*)(1 - P(\hat{s}_j^*))
\end{aligned}$$

as $v(s) < v(1) \leq 0$ for all $s \in S \setminus \{1\}$. For the second case, for all $j \in N$ unless $\hat{s}_j^* = 1$

$$\begin{aligned}
0 &= \int_1^1 (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s) ds \\
&< \int_{\hat{s}_j^*}^1 (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s) ds = v(\hat{s}_j^*)(1 - P(\hat{s}_j^*))
\end{aligned}$$

as $0 \leq v(0) < v(s)$ for all $s \in S \setminus \{0\}$.

We see that we must have that $\beta_i^* \in (0,1)$ for all $i \in N$. Rearranging the implied equality, we have that for any $i, j \in N$

$$\begin{aligned}
&\int_{\hat{s}_i^*}^1 (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s) ds = \int_{\hat{s}_j^*}^1 (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s) ds \\
&\Leftrightarrow \int_{\hat{s}_i^*}^{\hat{s}_j} (v(s) - \frac{1-P(s)}{p(s)} v'(s))\, p(s) ds = 0
\end{aligned}$$

which implies $\hat{s}_i^* = \hat{s}_j^*$ by IVV.

Proof of Proposition 5

For any fixed n, by Proposition 6 we have that $\hat{s}_i^{(n)} = \hat{s}^{(n)}$ for all $i \in N$ under the symmetric dimension space assumption. Given the principal's optimal thresholds $\hat{s}^{(n)}$, the saddle point problem in Theorem 1 is

$$\begin{aligned}
&\min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \overline{W}(\hat{s}^{(n)}, \beta) \\
&= \sum_{i \in N} [\beta_i \int_{\hat{s}^{(n)}}^1 (\frac{1}{n} v(s) - \frac{1-P(s)}{p(s)} \frac{1}{n} v'(s))\, p(s) ds + \frac{1}{n} w(1 - P(\hat{s}^{(n)}))]
\end{aligned}$$

This has solution $\beta_i = \frac{1}{n}$ for all $i \in N$. The optimal threshold $\hat{s}^n$ can then be characterized by the following variational inequality. For all $s \in S$ we must have

$$(s - \hat{s}^{(n)})(\frac{1}{n}(\frac{1}{n}v(s) - \frac{1-P(s)}{p(s)}\frac{1}{n}v'(s)) + \frac{1}{n}w) \geq 0$$

$$\Leftrightarrow$$

$$(s - \hat{s}^{(n)})((\frac{1}{n}v(s) - \frac{1-P(s)}{p(s)}\frac{1}{n}v'(s)) + w) \geq 0$$

If $w > 0$, there exists an $\underline{n}$ such that for all $\tilde{n} \geq \underline{n}$, since $v(s)$ is bounded as it is continuously differentiable with domain $[0,1]$.

$$\frac{1}{\tilde{n}}(v(0) - \frac{1-P(\hat{0})}{p(\hat{0})}v'(0)) + w > 0$$

By IVV, we then have that $(\frac{1}{n}v(s) - \frac{1-P(s)}{p(s)}\frac{1}{n}v'(s)) + w > 0$ for all $s \in S$. Thus the only solution to the variational inequality is $\hat{s}^{(\tilde{n})} = 0$, as if $\hat{s}^{(n)} > 0$ then the inequality is violated for all $s \in [0, \hat{s}^{(n)}]$.

We can make an analogous argument when $w < 0$ to show the second part.

Proof of Theorem 2

The same proof as in Theorem 1 applies up to Step 3. Step 3 still applies but the upper bound problem modified so that we no longer have $\overline{U}(0) \in \mathbb{R}^n_{\geq 0}$, but instead $\overline{U}(0) \in \{U \in \mathbb{R}^n : \sum_{i \in N} U_i = 0\} \equiv \mathcal{U}^{SN}$. The upper bound problem is then

$$\begin{aligned}
& \min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathcal{U}^{SN}} \overline{W}(\hat{s}, \overline{U}(0), \beta) \\
= & \sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathcal{U}^{SN}} \overline{W}(\hat{s}, \overline{U}(0), \beta^*) \\
= & \sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathcal{U}^{SN}} \min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \overline{W}(\hat{s}, \overline{U}(0), \beta)
\end{aligned}$$

with

$$\overline{W}(\hat{s}, \overline{U}(0), \beta) = \sum_{j \in N} [-\beta_j \overline{U}_j(0) + \int_{\hat{s}_i}^{1} (\phi_i(s) + w_i) \, p(s) ds]$$

The saddle point values β^* must be such that $\beta_i^* = \frac{1}{n}$ for all $i \in N$. Otherwise if $\beta_j^* > \frac{1}{n}$ for some $j \in N$ then for any $U < 0$, we can choose $\overline{U}_j(0) = U$, and for all $l \in N \setminus \{j\}$ $\overline{U}_l(0) = -\frac{1}{n-1} U$. This satisfies the sum-narrow participation constraint and allows us to obtain an arbitrarily large payoff by choosing U.

With $\beta^* = (\frac{1}{n}, ..., \frac{1}{n})$, we can attain the value of the upper bound in the full problem by setting $\overline{U}(0)$ such that $\sum_{i \in N} \overline{U}_i(0) = 0$ and for any $i, j \in N$ the statistical correctness constraint $\overline{U}_i(0) - v_i(s)(1 - P(\hat{s}_i)) = \overline{U}_j(0) - v_j(s)(1 - P(\hat{s}_j))$ holds. This can be achieved by

$$\overline{U}_i(0) = v_i(s)(1 - P(\hat{s}_i)) - \frac{1}{n} \sum_{j \in N} v_j(s)(1 - P(\hat{s}_j)) \text{ for all } i \in N$$

Proof of Proposition 7

For the second case where $v(s) \geq 0$ for all $s \in S$ and the principal is worse off under narrow inference, from Proposition 3 the result holds if the principal is restricted to implement a threshold strategy. Since the principal can also implement a non-threshold strategy in the rational benchmark, there is then an additional gain to the principal from rational inference over narrow inference without IVV.

For the first case, let g^{int} be an interval strategy with k intervals. Let $N_l \subseteq N$ be the subset of dimensions such that the agent takes the action $a_i = 1$ for some type in interval l; if $i \in N_l$ then $g_i^{int}(a_i|s) = 1$ for all $s \in [z_{l-1}, z_l)$.

For each dimension $i \in N$, we define the smallest type $\underline{s}_i$ that both takes the action $a_i = 1$ under g^{int} and for which the principal would get a positive payoff if all types above where to take the action $a_i = 1$; $\underline{s}_i = \min\{s \in S : g_i^{int}(1|s) = 1 \text{ and } w_i + v_i(s) \geq 0\}$. Each $\underline{s}_i$ is in one of the k intervals of g^{int}, for each $i \in N$ denote this interval by $[z_{l_i - 1}, z_{l_i})$. We have that $\underline{s}_i$ is well defined because the lower bound of any $[z_{l_i-1}, z_{l_i})$ is closed. The set of dimensions at which action 1 is taken in this interval is denoted N_{l_i} and includes i. Let $i^* =$

$\arg\min_{i\in N} v_i(\underline{s}_i)(1-P(\underline{s}_i))$.

The principal's welfare when implementing deterministic interval strategy g^{int} in the rational benchmark can be obtained from the expression (41).

$$\sum_{l=1}^{k}\sum_{i\in N_l}\left[(w_i+v_i(z_{l-1}))(1-P(z_{l-1}))-(w_i+v_i(z_l))(1-P(z_l))\right] \tag{42}$$

We can write the principal's payoff from action $a_i=1$ in interval $[z_{l-1},z_l)$ as

$$\begin{aligned}&(w_i+v_i(z_{l-1}))(1-P(z_{l-1}))-(w_i+v_i(z_l))(1-P(z_l))\\&=(w_i+v_i(z_{l-1}))(P(z_l)-P(z_{l-1}))-(v_i(z_l)-v_i(z_{l-1}))(1-P(z_l))\end{aligned} \tag{43}$$

and then rewrite the principal's welfare

$$\sum_{l=1}^{k}\sum_{i\in N_l}(w_i+v_i(z_{l-1}))(P(z_l)-P(z_{l-1}))-\sum_{l=1}^{k}(1-P(z_l))\sum_{i\in N_l}(v_i(z_l)-v_i(z_{l-1})) \tag{44}$$

We can write the second term in (44) as

$$\begin{aligned}&\sum_{l=1}^{k}(1-P(z_l))\sum_{j\in N_l}(v_j(z_l)-v_j(z_{l-1}))\\&=\sum_{l=1}^{k}(P(z_l)-P(z_{l-1}))\sum_{m=1}^{l}\sum_{j\in N_m}(v_j(z_m)-v_j(z_{m-1}))\end{aligned} \tag{45}$$

For g^{int} to be IC, the weak monotonicity condition in Lemma A.1 requires that for any intervals $l,m\in\{1,...,k\}$ and any types $s\in[z_{l-1},z_l)$, $s'\in[z_{m-1},z_m)$

$$\sum_{i\in N_l}(v_i(s)-v_i(s'))\geq\sum_{i\in N_m}(v_i(s)-v_i(s'))$$

From this and continuity of $v_i(.)$, we have that for any $l,h\in\{1,..,k\}$

$$\sum_{m=1}^{l}\sum_{i\in N_m}(v_i(z_m)-v_i(z_{l-m}))\geq\sum_{m=1}^{l}\sum_{j\in N_h}(v_j(z_m)-v_j(z_{m-1})) \tag{46}$$

In particular, this applies to the first interval that the action $a_{i^*}=1$ is taken;

$h = l_{i^*}$. Since for any $h \in \{1, ..., k\}$

$$\begin{aligned} &\sum_{l=1}^{k} (P(z_l) - P(z_{l-1})) \sum_{m=1}^{l} \sum_{j \in N_h} (v_j(z_m) - v_j(z_{m-1})) \\ &= \sum_{l=1}^{k} (1 - P(z_l)) \sum_{j \in N_h} (v_j(z_l) - v_j(z_{l-1})) \end{aligned} \tag{47}$$

combining (45), (46) and (47) then gives us

$$\sum_{l=1}^{k} (1 - P(z_l)) \sum_{j \in N_l} (v_j(z_l) - v_j(z_{l-1})) \geq \sum_{l=1}^{k} (1 - P(z_l)) \sum_{i \in N_{l_{i^*}}} (v_i(z_l) - v_i(z_{l-1}))$$

and thus from (44) the following upper bound on the welfare of the principal.

$$\sum_{l=1}^{k} \sum_{i \in N_l} (w_i + v_i(z_{l-1}))(P(z_l) - P(z_{l-1})) - \sum_{l=1}^{k} (1 - P(z_l)) \sum_{i \in N_{l_{i^*}}} (v_i(z_l) - v_i(z_{l-1})) \tag{48}$$

Since $v_i(.)$ is increasing and $v_i(s) \leq 0$ for all $s \in S$, this in turn is upper bounded by

$$\sum_{l=1}^{k} \sum_{i \in N_l} (w_i + v_i(z_{l-1}))(P(z_l) - P(z_{l-1})) - \sum_{l=l_{i^*}}^{k} (v_{i^*}(z_l) - v_{i^*}(z_{l-1}))(1 - P(z_l))$$

$$\leq \sum_{j \in N} w_j (1 - P(\underline{s}_j)) \tag{49}$$

$$+ \sum_{l=l_{i^*}}^{k} \sum_{i \in N_l} [v_i(z_{l-1})(P(z_l) - P(z_{l-1})) - (v_{i^*}(z_l) - v_{i^*}(z_{l-1}))(1 - P(z_l))] \tag{50}$$

$$\leq \sum_{j \in N} w_j (1 - P(\underline{s}_j)) + \int_{\underline{s}_{i^*}}^{1} (v_{i^*}(s) - \frac{1 - P(s)}{p(s)} v'_{i^*}(s))\, p(s)\, ds \tag{51}$$

Under narrow inference, the principal can achieve this upper bound by implementing a deterministic threshold strategy g^* such that $g_i^*(1|s) = \mathbb{1}\{s \geq \underline{s}_i\}$ for all $i \in N$. This gives the result.

The principal does this by implementing beliefs such that $\bar{t}_{i^*}(0) = 0$, $\bar{t}_{i^*}(1) =$

$-v_{i^*}(\underline{s}_{i^*})$, and for all $j \in N \setminus \{i^*\}$

$$\begin{aligned}
\overline{t}_j(0) &= v_j(\underline{s}_j)(1 - P(\underline{s}_j)) - v_{i^*}(\underline{s}_{i^*})(1 - P(\underline{s}_{i^*})) \geq 0 \\
\overline{t}_j(1) &= \overline{t}_j(0) - v_j(\underline{s}_j)
\end{aligned}$$

from Proposition 2 we can find a transfer function that implements these beliefs since g^* is a deterministic threshold strategy.

Proof of Theorem 3

Step 1 of Theorem 1 works as before. From Step 2 onwards, we replace the objective by

$$\widehat{\widehat{W}}(x, \overline{U}(0), \beta) = \sum_{j \in N} [-\beta_j \overline{U}_j(0) + \int_{x_j}^{1} (\widehat{\widehat{\phi}}_j(P^{-1}(u)) + w_j) du]$$

where $P^{-1}(.)$ is the strictly increasing quantile function for the type distribution as before. This is an upper bound on the original objective as by definition of the upper concave envelope, $\int_x^1 (\widehat{\widehat{\phi}}_i(P^{-1}(u)) du = \widehat{\widehat{\Phi}}_i(x) \geq \Phi_i(x) = \int_x^1 (\phi_i(P^{-1}(u)) du$ for all $i \in N$, $x \in [0, 1]$. Since by Lemma 3 only deterministic threshold strategies are NIC and all deterministic threshold strategies can be made NIC by some transfer function, the new objective remains an upper bound of the full problem even without increasing $\phi_i(.)$.

Since $\widehat{\widehat{\phi}}_i(.)$ is increasing in s the new objective is concave for fixed β, and we can make the same argument as we made for Theorem 1 using the minimax theorem to get

$$\begin{aligned}
&\min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \widehat{\widehat{W}}(\hat{s}, \overline{U}(0), \beta) \\
= &\sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \widehat{\widehat{W}}(\hat{s}, \overline{U}(0), \beta^*) \\
= &\sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \min_{\beta \in [0,1]^n, \sum_{i \in N} \beta_i = 1} \widehat{\widehat{W}}(\hat{s}, \overline{U}(0), \beta)
\end{aligned}$$

In $\widehat{\overline{W}}(\hat{s}, \overline{U}(0), \beta)$ the term for $\overline{U}(0)$ is separable, therefore $\hat{s}^*(\beta) \in \arg\max_{\hat{s} \in [0,1]^n} \widehat{\overline{W}}(\hat{s}, \overline{U}(0), \beta)$ for all $\overline{U}(0)$. Thus, these thresholds maximize the new objective for fixed β. Therefore, we must have $\sup_{\hat{s} \in [0,1]^n, \overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \widehat{\overline{W}}(\hat{s}, \overline{U}(0), \beta^*) = \sup_{\overline{U}(0) \in \mathbb{R}^n_{\geq 0}} \widehat{\overline{W}}(\hat{s}^*(\beta^*), \overline{U}(0), \beta^*)$.

We now apply the same argument as Myerson (1981). For any $i \in N$

$$\int_{\hat{s}_i}^1 (\phi_i(s) - \widehat{\overline{\phi}}_i(s))\, p(s) ds = \int_{P(\hat{s}_i)}^1 (\phi_i(P^{-1}(u)) - \widehat{\overline{\phi}}_i(P^{-1}(u))) du$$
$$= \Phi_i(P(\hat{s}_i)) - \widehat{\overline{\Phi}}_i(P(\hat{s}_i))$$

We can show that $\Phi_i(P(\hat{s}_i^*(\beta))) = \widehat{\overline{\Phi}}_i(P(\hat{s}_i^*(\beta)))$ for any β. When $\hat{s}_i^*(\beta) \in \{0, 1\}$ this is clear. If $\hat{s}_i^*(\beta) \in (0, 1)$ then $\beta_i \widehat{\overline{\phi}}_i(\hat{s}_i^*(\beta)) + w_i = 0$, and by definition $\hat{s}_i^*(\beta)$ is the smallest type satisfying this. Since $\widehat{\overline{\Phi}}_i(P(s)) > \Phi_i(P(s))$ only in intervals $s \in [\underline{s}, \overline{s})$ where $\widehat{\overline{\phi}}_i(s)$ is constant, at $\hat{s}_i^*(\beta)$ we must have $\Phi_i(P(\hat{s}_i^*(\beta))) = \widehat{\overline{\Phi}}_i(P(\hat{s}_i^*(\beta)))$ otherwise we can find a smaller threshold in the maximizing set.

For any β we can write the old objective as

$$\begin{aligned}
&\overline{W}(\hat{s}, \overline{U}(0), \beta) \\
&= \sum_{j \in N} [-\beta_j \overline{U}_j(0) + \int_{\hat{s}_j}^1 (\beta_j \phi_j(s) + w_j)\, p(s) ds] \\
&= \sum_{j \in N} [-\beta_j \overline{U}_j(0) + \int_{\hat{s}_j}^1 (\beta_j \widehat{\overline{\phi}}_j(s) + w_j)\, p(s) ds \\
&+ \beta_j \int_{\hat{s}_j}^1 (\phi_j(s) - \widehat{\overline{\phi}}_j(s))\, p(s) ds] \\
&= \sum_{j \in N} [-\beta_j \overline{U}_j(0) + \int_{\hat{s}_j}^1 (\beta_j \widehat{\overline{\phi}}_j(s) + w_j)\, p(s) ds \\
&+ \beta_j (\Phi_j(P(\hat{s}_j)) - \widehat{\overline{\Phi}}_j(P(\hat{s}_j)))]
\end{aligned}$$

At $\hat{s}^*(\beta)$, since $\Phi_i(P(\hat{s}_i^*(\beta))) = \widehat{\overline{\Phi}}_i(P(\hat{s}_i^*(\beta)))$ the value the old and new objectives are identical for any β; $\overline{W}(\hat{s}^*(\beta), \overline{U}(0), \beta) = \widehat{\overline{W}}(\hat{s}^*(\beta), \overline{U}(0), \beta)$. This means that for any $\hat{s} \in [0,1]$

$$\overline{W}(\hat{s}^*(\beta), \overline{U}(0), \beta) = \widehat{\overline{W}}(\hat{s}^*(\beta), \overline{U}(0), \beta) \geq \widehat{\overline{W}}(\hat{s}, \overline{U}(0), \beta) \geq \overline{W}(\hat{s}, \overline{U}(0), \beta)$$

which then implies

$$\begin{aligned}
&\sup_{\hat{s}\in[0,1], \overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \min_{\beta\in[0,1]^n, \sum_{i\in N}\beta_i=1} \overline{W}(\hat{s}, \overline{U}(0), \beta) \\
&= \sup_{\overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \min_{\beta\in[0,1]^n, \sum_{i\in N}\beta_i=1} \overline{W}(\hat{s}^*(\beta^*), \overline{U}(0), \beta) \\
&= \sup_{\overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \min_{\beta\in[0,1]^n: \sum_{i\in N}\beta_i=1} \widehat{\overline{W}}(\hat{s}^*(\beta^*), \overline{U}(0), \beta) \\
&= \min_{\beta\in[0,1]^n, \sum_{i\in N}\beta_i=1} \sup_{\hat{s}\in[0,1]^n, \overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \widehat{\overline{W}}(\hat{s}, \overline{U}(0), \beta) \\
&= \sup_{\hat{s}\in[0,1]^n, \overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \widehat{\overline{W}}(\hat{s}, \overline{U}(0), \beta^*) \\
&= \sup_{\overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \overline{W}(\hat{s}^*(\beta), \overline{U}(0), \beta^*) \\
&= \min_{\beta\in[0,1]^n: \sum_{i\in N}\beta_i=1} \sup_{\hat{s}\in[0,1]^n, \overline{U}(0)\in\mathbb{R}^n_{\geq 0}} \overline{W}(\hat{s}, \overline{U}(0), \beta)
\end{aligned}$$

giving us a minimax result for $\overline{W}(.)$ also. We can then apply Step 4 from Theorem 1 to show that this objective value can be achieved in the full problem, which completes the proof.

References

[1] Armstrong, M. (2016). Nonlinear pricing. *Annual Review of Economics 8*(1), 583–614. 25

[2] Bohren, J. A. and D. N. Hauser (2021). Learning with heterogeneous misspecified models: Characterization and robustness. *Econometrica 89*(6), 3025–3077. 9

[3] Camara, M. K. (2022). Computationally tractable choice. In *EC*, pp. 28. 9

[4] Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica 85*(2), 453–488. 15, 40

[5] Eisenhuth, R. (2019). Reference-dependent mechanism design. *Economic Theory Bulletin 7*(1), 77–103. 10

[6] Eliaz, K. and R. Spiegler (2006). Contracting with diversely naive agents. *The Review of Economic Studies 73*(3), 689–714. 10

[7] Eliaz, K. and R. Spiegler (2020). A model of competing narratives. *American Economic Review 110*(12), 3786–3816. 9

[8] Eliaz, K. and R. Spiegler (2024). News media as suppliers of narratives (and information). *arXiv preprint arXiv:2403.09155*. 9

[9] Enke, B. and F. Zimmermann (2019). Correlation neglect in belief formation. *The review of economic studies 86*(1), 313–332. 8

[10] Esponda, I. and D. Pouzo (2016). Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica 84*(3), 1093–1130. 9

[11] Eyster, E. and M. Rabin (2005). Cursed equilibrium. *Econometrica 73*(5), 1623–1672. 10

[12] Farhi, E. and X. Gabaix (2020). Optimal taxation with behavioral agents. *American Economic Review 110*(1), 298–336. 10

[13] Feldman, N. E., P. Katuš čÁk, and L. Kawano (2016). Taxpayer confusion: Evidence from the child tax credit. *American Economic Review 106*(3), 807–835. 11

[14] Fernbach, P. M., A. Darlow, and S. A. Sloman (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science 21*(3), 329–336. 8

[15] Frick, M., R. Iijima, and Y. Ishii (2020). Misinterpreting others and the fragility of social learning. *Econometrica 88*(6), 2281–2328. 9

[16] Fudenberg, D., G. Lanzani, and P. Strack (2021). Limit points of endogenous misspecified learning. *Econometrica 89*(3), 1065–1098. 9

[17] Graeber, T. (2023). Inattentive inference. *Journal of the European Economic Association 21*(2), 560–592. 9

[18] He, S. and S. Kučinskas (2024). Expectation formation with correlated variables. *The Economic Journal 134*(660), 1517–1544. 8

[19] Heidhues, P. and B. Kőszegi (2010). Exploiting naivete about self-control in the credit market. *American Economic Review 100*(5), 2279–2303. 10

[20] Heidhues, P., B. Kőszegi, and P. Strack (2018). Unrealistic expectations and misguided learning. *Econometrica 86*(4), 1159–1214. 9

[21] Herweg, F., D. Müller, and P. Weinschenk (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review 100*(5), 2451–2477. 10

[22] Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory 123*(2), 81–104. 10, 37

[23] Jehiel, P. (2011). Manipulative auction design. *Theoretical economics 6*(2), 185–217. 10

[24] Jehiel, P. and K. Mierendorff (2024). Auction design with data-driven misspecifications: Inefficiency in private value auctions with correlation. *Theoretical Economics*, (Forthcoming). 10

[25] Kőszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature 52*(4), 1075–1118. 10

[26] Kőszegi, B. and F. Matějka (2020). Choice simplification: A theory of mental budgeting and naive diversification. *The Quarterly Journal of Economics 135*(2), 1153–1207. 9

[27] Lian, C. (2021). A theory of narrow thinking. *The Review of Economic Studies 88*(5), 2344–2374. 9

[28] Liebman, J. B. and R. J. Zeckhauser (2004). Schmeduling. 11

[29] Lockwood, B. B. (2020). Optimal income taxation with present bias. *American Economic Journal: Economic Policy 12*(4), 298–327. 10

[30] Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601. 39, 42

[31] Mirman, L. J. and D. Sibley (1980). Optimal nonlinear prices for multiproduct monopolies. *The Bell Journal of Economics 11*(2), 659–670. 15

[32] Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research 6*(1), 58–73. 35, 60

[33] O'Donoghue, T. and M. Rabin (2006). Optimal sin taxes. *Journal of Public Economics 90*(10-11), 1825–1849. 10

[34] Pearl, J. (2009). *Causality*. Cambridge university press. 7

[35] Piccione, M. and A. Rubinstein (2003). Modeling the economic interaction of agents with diverse abilities to recognize equilibrium patterns. *Journal of the European economic association 1*(1), 212–223. 10

[36] Rabin, M. and G. Weizsäcker (2009). Narrow bracketing and dominated choices. *American Economic Review 99*(4), 1508–1543. 9

[37] Read, D., G. Loewenstein, M. Rabin, G. Keren, and D. Laibson (1999). Choice bracketing. *Journal of Risk and Uncertainty 19*(1/3), 171–197. 9

[38] Rees-Jones, A. and D. Taubinsky (2020). Measuring "schmeduling". *The Review of Economic Studies 87*(5), 2399–2438. 11

[39] Rubinstein, A. (1993). On price recognition and computational complexity in a monopolistic model. *Journal of Political Economy 101*(3), 473–484. 10

[40] Schumacher, H. and H. C. Thysen (2022). Equilibrium contracts and boundedly rational expectations. *Theoretical Economics 17*(1), 371–414. 9

[41] Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics 8*(1), 171 – 176. 51

[42] Spiegler, R. (2016). Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics 131*(3), 1243–1290. 9

[43] Spiegler, R. (2020). Behavioral implications of causal misperceptions. *Annual Review of Economics 12*(1), 81–106. 20

[44] Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association 13*(1), 130–167. 10

[45] Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science 4*(3), 199–214. 9

[46] Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making 12*(3), 183–206. 9

[47] Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *science 211*(4481), 453–458. 9

[48] Vorjohann, P. (2024). Reference-dependent choice bracketing. *Available at SSRN: https://ssrn.com/abstract=4927177*. 9