Narrow Inference and Incentive Design*

Alexander Clyde†

This version: January 30, 2026

First version: October 16, 2024

[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 sort different types of agents into choosing different action combinations. The agents make choices on multiple dimensions, and infer the effect of each action separately without properly controlling for the other actions. I fully characterize the principal's optimal mechanism when facing agents who do such 'narrow' inference, and contrast it with their optimal mechanism when the agents are fully rational.

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, Dajana Xhani and audiences at Humboldt-Universität zu Berlin, Bilkent University, Edinburgh University, the Midwest Theory Conference at Virginia Tech, the Barcelona Summer Forum Workshop on Preferences, Bounded Rationality, and Strategic Interactions, the Conference on Economic Design at University of Essex, the Nordic Economic Theory Workshop at University of Copenhagen, the SAET Conference in Ischia and the Stony Brook International Game Theory Conference for helpful questions and comments. I gratefully acknowledge financial support from an Economic and Social Research Council studentship (1923039), from Leverhulme Trust grant RPG-2023-120 and from ERC Grant 101163162.

†Department of Economics, Aalto University School of Business; Address: Ekonominaukio 1, FI-02150 Espoo, Finland; Email: alexander.clyde.econ@gmail.com or alexander.clyde@aalto.fi

1 Introduction

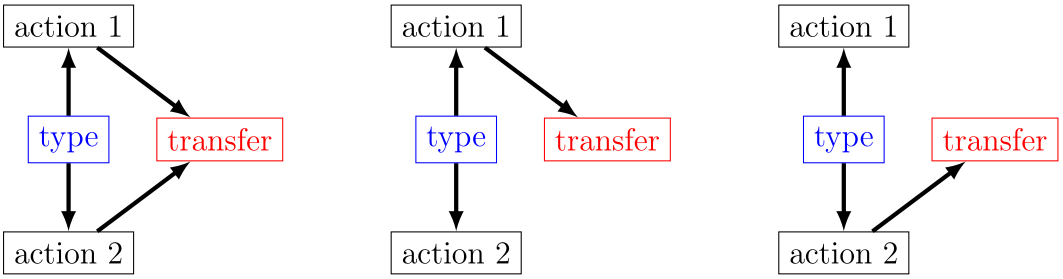
Understanding the incentives we face requires understanding what the causal consequences of our actions are on outcomes that we care about. Moreover, many incentive structures are complex, multidimensional and opaque, with the final consequences uncertain and unrealized at the point decisions are taken. For example, workers have to form beliefs about how their choices of effort, occupation and education affect their ultimate earnings. In such circumstances, economic models often implicitly assume that people can form beliefs about the incentives they face as if they are making sophisticated causal inference from any available data. However, work in experimental and behavioural economics suggests that people struggle with various aspects of performing causal inference.[1]

Since causal understanding matters for the perception of incentives, when designing incentives it is important to take into account people's limited ability to perform causal inference. In this paper, I provide a full characterization of how a principal should design incentives for agents who form beliefs about the effects of their actions according to a procedure I call *narrow inference*. This model of belief formation is able to link models of causal misperceptions from the literature on misspecified models with work in behavioural economics on narrow choice bracketing.

To understand narrow inference, consider a large organization that is determining its wage structure. A worker in the organization has to decide whether to make human capital investments in technical skills and/or managerial skills. Under narrow inference, the worker estimates the earnings benefits of acquiring each skill narrowly and separately, by comparing the average wage of workers in the organization who have the skill in question with the average wage of workers without that skill. 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 from obtaining different forms of human capital. The extent of this misperception affects how the organization wants to design their wage structure. Figure 1 illustrates this bias using the Directed Acyclic Graph (DAG) notation of Pearl (2009).

In this paper I analyze such incentive design problems. I explore how a princi-

[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.



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

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

pal would design an incentive mechanism if they knew the agents had this form of bounded rationality. I obtain a result characterizing the principal's optimal mechanism with agents who make narrow inference as the solution to a zero-sum game. This result is useful both because it provides a recipe for how to solve particular parameterized applications of an incentive design problem under narrow inference, and because it allows us to make clear comparisons between the principal's optimal mechanism under narrow inference with the principal's optimal mechanism in the rational benchmark.

I build on the characterization result to show how narrow inference changes the shape of the optimal mechanism. I show that this fundamentally depends on the nature of the contracting relation between the principal and the agents. If the principal is an employer who benefits from the agents taking costly actions, then the principal unambiguously gains from narrow inference and wants to implement that a greater proportion of agents take the actions. In contrast, if the principal is themselves a worker providing costly services benefiting the agent, the opposite conclusions hold. Finally, I show that under narrow inference the principal can sometimes benefit from preventing the agent from taking some actions separately, a form of 'pure bundling', something that has no benefit to the principal in the rational benchmark.

In the model, the agents face a binary decision problem on whether to take an action or not on multiple dimensions. The principal chooses a function mapping the agents' actions to a 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 utility of the agents can be decomposed into a type-dependent predictable part that they fully understand, and a transfer part that they need to infer. Each agent's type is private information, and the principal's choice of transfer function screens different types of agents into choosing different actions. With fully rational agents who know and understand the true transfer function this is a standard monopolistic screening problem.

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 is correlated across dimensions, something that is possible due to joint dependence on the type variable.

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

Example 1. The agents have two dimensions of action, whether to obtain technical skills $a_1 = 1$ or not $a_1 = 0$ and whether to obtain managerial skills $a_2 = 1$ or not $a_2 = 0$. The principal, an employer, sets an earnings schedule for roles within their organization 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 agents' 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 suffice for all incentive constraints to hold.

Now consider what happens if the agents use 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 managerial skills mixes the earnings of those who get technical skills 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 agents perceive allows the principal to implement the same action choices for each type while reducing earnings across the type distribution. $\triangle$

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 state results adapted from Carroll (2017) which describe the principal's optimal mechanism for the rational benchmark. Under a standard regularity assumption, the principal's optimal mechanism is fully separable across dimensions. Facing such a mechanism, the agents choose 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 agents' 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 demonstrates that narrow inference amplifies the perceived size of transfers relative to their actual size.

Thus, the size of the transfers required to implement any perceived size of incentives is reduced under narrow inference. This means that how welfare of the principal and the equilibrium strategy of the agents change under narrow inference relative to the rational benchmark depends on whether the agents are a recipient or a contributor of transfers in equilibrium. I show that there are two different cases where the conclusions are unambiguous. When the agents take actions that have predictable utility costs to themselves, but benefit the principal, the agents are recipients of transfers. In this case the principal benefits from narrow inference, and implements that a greater fraction of different types of agent take the costly actions on any dimension. On the other hand, when agents' actions benefit themselves but are costly to the principal the opposite conclusions hold.[2]

[2]Whether the actions have positive or negative predictable utility for the agents is not an arbitrary re-labelling of the actions, as we normalize the predictable utility from the zero action

The first case is in line with the principal employing the worker, where human capital investments have acquisition costs for the worker but increase profitability for the employing organization. In the second case, the contractual relationship is flipped, with the principal now a worker offering services to the agents who assess the cost of each dimension of service provided using narrow inference. In both applications the final transfer is undetermined at the point when the agents makes their choices but there is ample data available on how past choices of similar agents vary with wages and prices from which the agents can form estimates. For example, consider online platforms like Glassdoor, LinkedIn or Indeed which provided information on human capital and salaries and help match employees to employers.

I then consider what happens if actions can be bundled, so that certain action are linked and the agents are forced to take them jointly if at all. When actions are costly to the agents but benefit the principal this does not improve the welfare of the principal. However in the setting where the principal offers costly services for the agents' benefit, bundling can have significant gains for the principal. This is because it mitigates the negative effects of narrow inference on the principal's welfare by reducing the extent of the distortion in the agents' beliefs. In the rational benchmark, bundling has no advantages to the principal as it only restricts the set of actions the principal can induce the agents to take.

Finally, I discuss the implications of the results concerning principal welfare and pure bundling for applications of the model to employment contracting. The results suggest that large organizations employing many agents may want to present the 'choice architecture' of the employment relationship as consisting of separate tasks or jobs so as to exploit narrow inference. In contrast, a self-employed worker would want to present different work tasks in a joint and connected way to potential buyers of their services. I present some evidence suggestive of this pattern.

The paper takes a step towards understanding how errors in causal inference might affect how we should design economic incentives by studying a novel model of belief formation. In Section 5.2, I discuss the key assumptions of the model and how they could be relaxed in future work building on this paper.

to zero. Without the normalization the result would be stated in terms of the difference in predictable utilities between the two actions.

Literature Review

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 that can resolve tensions between narrow decision making and global budget balance.

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). Frick et al. (2022) propose a related concept, Assortativity Neglect Equilibrium, that models incorrect inference applied to data observed from the rest of the population. This is similar to narrow inference, which is an inferential error by agents using data on the population of agents interacting with the mechanism. 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.[3]. The first papers to explicitly apply the ABEE concept to design problems are Jehiel (2011) and Jehiel and Mierendorff (2024) who study auction design where bidders receive limited feedback. A paper explicitly studying narrow bracketing and auction design is Eisenhuth (2019), who uses a prospect theoretic formulation of narrow bracketing.

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).

[3]The behaviour of the principal and the agents under narrow inference can be formulated as an ABEE of an extensive form game, and I discuss this in more detail in additional Appendix Section A.1

2 Model

Each agent faces a multidimensional decision. Let $A = \{0,1\}^n$ be the agents' set of feasible action profiles. I refer to $i \in \{1,\dots,n\} \equiv N$ as a dimension, such that a_i is the agents' action in dimension i. An 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 $P(s) = \int_0^s p(\tilde{s})d\tilde{s}$. This type is the agents' private information and is unknown to the principal.

If the dimension i action a_i is taken by a type s agent it generates predictable utilities $v_i(s)a_i$ and $w_i(s)a_i$ for the agent and the principal respectively. The functions $v_i(s)$ and $w_i(s)$ are weakly increasing and continuously differentiable. In addition to the predictable utilities, 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 transfer $t \in \mathbb{R}$ represents a zero-sum transfer of surplus between the agent and the principal. The principal's payoff given that actions a are taken by a type s agent, resulting in transfer $t \in \mathbb{R}$ is

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

Throughout the paper, I make the following standard regularity assumption on the type distribution.[4]

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).

[4]In Section A.5 of the additional Appendix, I explore the implications of relaxing this assumption.

2.1 Mechanisms

I focus on a natural class of indirect mechanisms. Before the agents take any actions, the principal commits to a mechanism, which consists of a function mapping actions to transfers $t : A \to \mathbb{R}$.[5] After learning their type, each 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

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

I discuss this restriction on the allowable mechanisms in Section 5.2. 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 each 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 ($w_i(s) \geq 0$) or negative ($w_i(s) \leq 0$).

Take the applications suggested in the introduction. In the first, the agents are workers and the principal a large employing organization. Here $v_i(s) < 0$ is the cost of obtaining human capital and $w_i(s) > 0$ is the benefit to the organization of the human capital the worker acquires. The organization then chooses a earnings schedule t, which depends on the skills the workers acquire.

In the second application, the principal is the worker and the agents are organizations contracting for the worker's services. The principal offers to perform different tasks from which the agents derive positive utility $v_i(s) > 0$, but these

[5] We can incorporate uncertainty into the transfer function through a random additional variable $\omega \in \Omega \subseteq \mathbb{R}$, which is finite with distribution μ. The transfer function is then $t : A \times \Omega \to \mathbb{R}$. We can perform all the analysis of this paper with the expected transfer function $t(a) = \sum_{\omega \in \Omega} t(a, \omega)\mu(\omega)$, since the transfer enters linearly in the utilities of the principal and the agents.

tasks are costly for the principal to perform $w_i(s) < 0$. The principal sets a pricing schedule t for the services they are providing. The agents' type affects both their own utility from the tasks and the principal's costs of undertaking them.

The model best fits a population interpretation, where the principal chooses a mechanism and then repeatedly interacts with a population of short-run agents with types distributed according to $P(s)$. The equilibrium under both narrow inference and the rational benchmark is then a steady state of this process.

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 $a' \in A$

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

An agent always has the option of not participating in the mechanism, taking the actions $a = 0$ and obtaining zero transfer. Thus the transfer conditional on action combination $a = 0$ must be at least zero.[6]

$$t(0, ., 0) \geq 0 \tag{6}$$

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 interchange-

[6]Together with the rational incentive constraints this implies that for all $s \in S$, $U(s) \geq 0$

ably throughout the paper.

An 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 consider two different forms of participation constraint. The first are dimension-by-dimension *narrow participation constraints*. Under these constraints, for each dimension the expected transfer conditional on the action zero being taken on that dimension must be nonnegative. Formally, for all $i \in N$

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

The second form of participation constraint we consider is a *joint participation constraint*. Here the sum of expected transfers from action zero across dimensions must be non-negative.[7]

$$\sum_{i \in N} \bar{t}_i(0) \geq 0 \tag{13}$$

Note that the joint participation constraint is weaker than the narrow participation constraints. If all the narrow participation constraints hold then the joint participation constraint holds, but the converse is not true.

2.5 Interpretation of participation constraints

The narrow participation constraints reflect the idea that agents believe they can make a separate decision on participation on each dimension, so that if $\bar{t}_j(0) < 0$ for some dimension j they believe they can reject the transfer effect of dimension j and obtain zero transfer for that dimension — while continuing to receive transfer from the mechanism on other dimensions. Thus it is as if an agent perceives the transfer function is additively separable $\hat{t}(\tilde{a}) = \sum_{i \in N} \hat{t}_i(\tilde{a}_i)$, with three actions on each dimension $\tilde{a}_i \in \{NP, 0, 1\}$, where NP represents non-participation on that dimension. Agents then estimate that for each dimension $i \in N$ that $\hat{t}_i(NP) = 0$, $\hat{t}_i(0) = \bar{t}_i(0)$ and $\hat{t}_i(1) = \bar{t}_i(1)$.

Under the joint participation constraint, agents perceive that the decision to participate is all-or-nothing. Here an agent perceives the payoff from participation as $\sum_{i \in N} \bar{t}_i(0)$ and that taking the action $a_i = 1$ adds or subtracts the increment $\bar{t}_i(1) - \bar{t}_i(0)$ from this participation payoff. Any mechanism that satisfies the narrow participation constraints also satisfies this joint participation constraint, and we show how modified versions of the same results hold with both forms of participation constraints.

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 from Carroll (2017)

[7] Similarly to in the rational benchmark, these participation constraints can be combined with the narrow incentive constraints and written in terms of the narrow perceived utilities. The narrow participation constraints the require that for all $i \in N$ and $s \in S$, $\overline{U}_i(s) \geq 0$. The joint participation constraint requires that for all $s \in S$, $\sum_{i \in N} \overline{U}_i(s) \geq 0$.

adapted to our setting.[8]

It will be shown that the agents' strategy under the principal's optimal mechanism with rational agents takes a *threshold form* where there is a 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.*

$$\max_{\hat{s} \in [0,1]^n} W(\hat{s}) = \sum_{i \in N} \int_{\hat{s}_i}^{1} (\phi_i(s) + w_i(s))\, p(s)\, ds \tag{14}$$

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

$$t(a_1, ..., a_n) = \sum_{i \in N} t^i(a_i) \tag{15}$$

$$t^i(0) = 0, t^i(1) = -v_i(\hat{s}_i) \text{ for all } i \in N \tag{16}$$

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.[9]

[8] 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.

[9] We see in Section A.5 of the additional Appendix 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

Under narrow participation constraints, 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 agents' 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. With the joint participation constraint, the result is simpler and the shrinkage factors are equalized across dimensions so that $\beta_i = \frac{1}{n}$ for all $i \in N$.

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*

$$\min_{\beta\in[0,1]^n:\sum_{i\in N}\beta_i=1}\ \max_{\hat{s}\in[0,1]^n} \overline{W}(\hat{s};\beta) = \max_{\hat{s}\in[0,1]^n}\ \min_{\beta\in[0,1]^n:\sum_{i\in N}\beta_i=1} \overline{W}(\hat{s};\beta) \qquad (17)$$

with the value of the principal's objective given by

$$\overline{W}(\hat{s};\beta) = \sum_{i \in N} \int_{\hat{s}_i}^{1} (\beta_i \phi_i(s) + w_i(s))\, p(s)\, ds \tag{18}$$

Under the ***joint participation constraint****, the result holds with fixed weights* $\beta_i = \frac{1}{n}$ *for all* $i \in N$.

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 $w_i(s) = w(s)$ 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 halve the size of 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 agents are 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 show how under narrow inference pure bundling can be used to improve the principal's welfare in some cases, even though it can never benefit the principal in the rational benchmark of this model. I now present some preliminaries that are used in obtaining the characterization.

4.2 Preliminaries for Characterization Result

I first consider which beliefs can be induced by some transfer function. I show that beliefs must satisfy a statistical-correctness constraint, and that for any threshold strategy there is a valid additively separable transfer function that induces beliefs satisfying this statistical correctness constraint.

It will be useful in characterizing the principal's optimal mechanism under NIC to show how beliefs and the transfer function relate given 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 Directed Acyclic Graphs (DAGs) (Spiegler, 2020).

Lemma 1 (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*

$$\sum_{a_i \in A_i} g_i(a_i)\bar{t}_i(a_i) = \sum_{a_j \in A_j} g_j(a_j)\bar{t}_j(a_j) \tag{19}$$

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

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

□

For a fixed strategy and action distribution, this statistical correctness is necessary but not sufficient for a transfer function to exist that induces given beliefs. 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 agents takes a threshold form. I now show that any NIC strategy must take 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.

Lemma 2. *Every NIC strategy takes a* ***threshold form*** *for almost all $s \in S$.*

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

$$\bar{t}_i(1) = \bar{t}_i(0) - v_i(\hat{s}_i) \tag{20}$$

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 $\hat{U}_i(s,1) \geq \hat{U}_i(s,0)$ if and only if $s \geq \hat{s}_i$, in which case type s will choose $a_i = 1$. Given a threshold strategy, the beliefs in the proposition statement ensure that agents are indifferent between taking either action at the threshold. □

The next result shows how any 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 agents are choosing actions according to a threshold strategy.

Proposition 2. *For any 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*

$$t(1, \tilde{a}_{-i}) - t(0, \tilde{a}_{-i}) = t(1, a_{-i}) - t(0, a_{-i}) \tag{21}$$

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 □

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. Proposition 2 means that if the principal implements a threshold strategy, they have no payoff gain from implementing a transfer function that is not additive. Thus, the principal does not exploit the agents' potentially 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 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 agents' 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, using Proposition 2, we write both the principal's objective and the statistical correctness constraint from Lemma 1 only in terms of the threshold strategy and the narrow perceived transfer of the type taking action zero on one of the dimensions; $\overline{t}_i(0)$. We then show we can obtain an upper bound to the principal's problem that takes a max-min form, with the β weights in the proof coming from a rewriting of the Lagrange multipliers from the statistical correctness constraint. Under IVV[10], we can apply a standard minimax theorem argument to obtain a saddle point, allowing us to interchange the min and the max. Finally, we show that this minimax upper bound can be attained in the principal's full problem. Under the weaker joint participation constraint, the shrinkage weights must be equalized across dimensions and the proof can be modified to work in this case.

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.[11]

Proposition 3. *Assume the IVV assumption holds. Under both the narrow and joint participation constraints*

[10]Without IVV, we can obtain a modified version of the result by using the ironing technique of Myerson (1981). Details of how this works can be found in Online Appendix Section A.5

[11]In Section A.4 in the additional Appendix, we also study the effect of narrow inference on the average welfare of agents. The effect is generally ambiguous, and can be either positive or negative under all the conditions in this section.

1. *If 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 agents are narrow compared to the rational benchmark.*

2. *If 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 agents are narrow.*

Proof. In Appendix □

The intuition for this result is as follows. When the agents' action is costly in terms of the agents' 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 predictable utility benefits to the agent then the principal is a net recipient of transfers from the agent and the logic is reversed, with the principal able to extract less transfer from higher types of agent under narrow inference.[12]

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 predictable benefits to the principal for all types, 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 when actions have predictable costs to the principal for all types of agent.

Proposition 4. *Assume the IVV assumption holds. Under both the narrow and joint participation constraints*

1. *If for every $i \in N$ we have $w_i(s) > 0$ for all $s \in S$, then on each dimension the objective-maximizing thresholds are weakly lower with narrow agents than*

[12] In Section A.5 of the additional Appendix, Proposition 8 shows that Proposition 3 extends to when we don't have IVV.

under the rational benchmark, so $a_i = 1$ is taken by a larger proportion of types for all $i \in N$.

2. *If for every $i \in N$ we have $w_i(s) < 0$ 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 $\square$

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 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 the principal has a lower marginal benefit — in terms of transfer received — from a higher proportion of agents taking the action, but the same cost. The principal then wants to reduce the proportion of agents taking the actions on all dimensions.

Suppose that whether the agents do narrow or rational 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 is that the agents do narrow inference, the principal can try to educate the agents on how to do rational inference. Proposition 3 then shows in what cases the principal would gain from de-biasing the agents and in what cases they would not. Employment arrangements where workers have distinct contracts for different jobs, tasks or skills could be seen as creating a 'choice architecture' designed to exploit narrow inference. In contrast, arrangements where tasks or services rendered are clearly linked as being part of the same contractual relationship — so that agents make comparisons directly between different action combinations — could reduce or eliminate narrow inference.

This manipulation of the choice architecture could be seen as a form of 'bundling' in the way choices are presented to the agent. This presentational bundling is motivated by the possibility of mitigating the costs of the agents' bounded rationality rather that the more typical motive of exploiting the principal's monopoly power.[13] In contrast to this soft presentational bundling, in the

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

next section I explore when the principal can benefit from pure bundling — that is explicitly restricting the ability of agents to take certain actions separately.

4.4 Pure Bundling Under Narrow Inference

I now explore whether the principal can benefit from linking or bundling the different choices of the agents together. Pure bundling means that the principal can ensure that if an agent takes a given action then they also have to take some subset of actions on other dimensions.[14] I assume the agents perceive linked choices as joint, that is if the principal restricts the agents to having to choose action a_1 and a_2 together the agents do not perform narrow inference for each dimension separately but instead treats the linked action $a_{1\wedge 2} = a_1 \cdot a_2$ as being one decision from which they receive predictable utility $(v_1(s) + v_2(s)) \cdot a_{1\wedge 2}$ and infer narrowly the transfer $\bar{t}_{1\wedge 2}(a_{1\wedge 2})$.[15]

Under the rational benchmark, the ability of the principal to implement pure bundle through directly restricting the action choices available to the agents is irrelevant as they could achieve the same effect through penalizing the action choices through the transfer function. Under narrow inference the agents perceive the transfer function as being additive, so do not respond to interactions in the transfer function required to bundle action choices.[16]

Under the rational benchmark, bundling has no benefit to the principal. The principal is always able to achieve any bundling through the transfer function rather than directly restricting actions, and from Proposition 1 under IVV the principal's optimal mechanism is additively-separable and so features what Adams and Yellen (1976) call 'pure-components pricing'. In contrast, we show that under narrow inference pure bundling can help mitigate against losses to the principal caused by narrow inference.

[14] This is in line with Adams and Yellen (1976) definition of pure bundling, in contrast to mixed bundling where the transfer function is not additively-separable but does not prohibit actions being taken separately.

[15] Without this assumption the agent would be forced to take actions that they didn't realize they were choosing. For example if actions 1 and 2 were bundled and they chose action 1 as under narrow inference for that dimension alone, then they would be choosing action 2 but without the predictable utility from action 2 entering in their utility.

[16] For example, if the principal wants to ensure that action $a_1 = 1$ is never taken separately from $a_2 = 1$ through the transfer function then $t(a_1 = 1, a_2 = 0) = \infty$ and $t(a_1 = 0, a_2 = 1) = \infty$. In the rational benchmark this always deters agents from taking these action combinations but under narrow inference if there are some agents choosing $(a_1, a_2) = (1, 1)$ and none taking the actions separately then agents' perceive the transfer from each action as $\bar{t}_1(a_1) = \bar{t}_2(1) = t(1, 1) < \infty$ and so may not be deterred from taking the actions separately in equilibrium.

Formally we model pure bundling through a reduction of the space of dimensions into a merged one. We say that a dimension space M is a *merge* of $N \supset M$ if there is a partition of N according to the dimensions in M where for each $i \in M$, there is a partition cell $\mathcal{P}(i)$ so that $v_i(s) = \sum_{j \in \mathcal{P}(i)} v_j(s)$ and $w_i(s) = \sum_{j \in \mathcal{P}(i)} w_j(s)$. Thus any partition cell $\mathcal{P}(i)$ represents a bundle of actions from the full dimension space that the agent must take jointly if at all under pure bundling. Comparing the principal's best-case mechanism in the merged and full dimension spaces then allows us to see if they can benefit from pure bundling.

Proposition 5. *Assume the IVV assumption holds and the agents do narrow inference.*

1. *If for every $i \in N$ we have $v_i(s) \leq 0$ for all $s \in S$, the principal cannot achieve a strictly higher payoff under a merge of the dimension space $M \subset N$ under both narrow and joint participation constraints.*

2. *If for every $i \in N$ we have $w_i(s) < 0$ and $v_i(s) \geq 0$ for all $s \in S$, the principal can always obtain a weakly higher payoff under a merge of the dimension space $M \subset N$ under narrow participation constraints.*

Proof. In Appendix □

Pure bundling reduces or eliminates the distortions in agents perceptions of the transfer under narrow inference, and also restricts the set of strategies the principal can implement on the bundled dimensions. When the principal is a net contributor of transfers in equilibrium then neither of these channels is advantageous to them. However, in cases where the principal is a recipient of transfers in equilibrium, narrow inference causes agents to over-perceive the size of the transfers. Pure bundling can benefit the principal by increasing the average transfer they extract from the agents for any fixed thresholds.

Under narrow participation constraints we can show that for any subset of bundled action dimensions, the principal increases the expected transfer they receive by preventing the types who did not take all of the actions from taking any of the actions in the subset. When these actions are costly to the principal, this unambiguously improves the principal's payoff. With a joint participation constraint, this is no longer true and the principal can both benefit or lose out from pure bundling depending on the size of the loss from having to implement the same threshold for the bundled dimensions.[17]

[17] For example, take the case with two dimensions and $v_i(s) = rs$, $w_i(s) = -c_i$ with $r, c_i > 0$

5 Discussion

5.1 Narrow Inference and the Choice Architecture of Employment

The analysis of this paper suggests that how narrow inference affects incentive design depends fundamentally on the bargaining relationship between the parties involved. When the principal is a large organization with enough bargaining power to design a wage structure for their workers, the principal prefers some degree of separation and unbundling of worker tasks in order to exploit the workers' narrow inference. When the principal is themselves a worker with scarce skills who has enough bargaining power to set the prices at which they will sell their services, the worker-principal prefers choice architecture presenting decisions as joint — and can even benefit from pure bundling that explicitly restricts certain action combinations being taken separately.

There are some empirical patterns observed in labour markets that fit this taxonomy. Employment arrangements featuring remuneration that is separate for different tasks over which the workers have some degree of autonomy is commonplace, with examples including pay for physicians (Dumont et al., 2008) and academics (Brickley and Zimmerman, 2001). Gig-economy platforms where workers are paid for separate tasks performed whilst working for the same organization are economically significant,[18] and represent a clear case where the choice architecture facing workers presents different actions separately. The total compensation such gig workers receive is often low compared to comparable workers (Mishel, 2018), in line with Theorem 1 where narrow inference reduces the transfers required to implement given action combinations.

In these settings workers face large employers who have substantial bargaining power. In contrast, workers who have greater control over their contractual terms often prefer to present jobs as a clearly defined set of tasks over which joint

for $i \in \{1, 2\}$. If $c_1 + c_2 \geq 2r$, $r > 2c_1$ and $r < 2c_2$ then under narrow inference the principal's optimal threshold when the two dimensions are bundled is $\hat{s} = 1$ giving them zero payoff. When the dimensions are not bundled the principal sets thresholds $\hat{s}_1 = \frac{1}{2} + \frac{c_1}{2}$ and $\hat{s}_2 = 1$ under narrow inference with joint participation constraint which gives them a positive payoff. When the principal faces narrow participation constraints, the principal sets thresholds $\hat{s}_1 = \hat{s}_2 = 1$ in the unbundled case also and so receives zero payoff, reflecting part 2 of Proposition 5. Note that when $2c_1 = 2c_2 < r$ the principal gains from pure bundling under the joint participation constraint also.

[18] Gallup (2018)'s survey finds that 36% of US workers have a gig-economy job according to a broad definition.

decisions are made. Occupational licensing restricts certain tasks as having to be performed together. It is associated with higher worker compensation (Kleiner and Krueger, 2013) and is often worker initiated.[19] Similarly, trade unions often engage in jurisdictional disputes asserting their members as having exclusive rights to perform certain bundles of tasks.[20]

Narrow Inference provides an explanation for the differences in the choice architecture workers face across these settings based on worker's limited ability to understand the effects of different choice dimensions on total compensation. This can be complementary with an explanation based on task complementarity or artificial entry barriers.

5.2 Model Assumptions and Possible Extensions

Separable Predictable Costs and Benefits

I assume the costs and benefits of the actions are additive and have no interactions across dimensions. This fits more cleanly with narrow inference, the agents perceiving complementarity or substitutability in the non-transfer part of their utility seems inconsistent with treating them as separate decision problems. A possible generalization would be to allow narrow inference misperceptions to extend to non-transfer components of the agents' payoff. Allowing interactions only in the principal's predictable utilities would also be consistent with narrow inference. However, this would mean the result from Proposition 1 that the principal's optimal mechanism takes a simple threshold form would not hold unless we make a more complicated assumption on the primitives than increasing virtual values (IVV).

Single Type Dimension

The assumption of a single dimension of type fits an interpretation that there is a single latent factor such as ability or company size affecting the agents utilities. It is possible to express the primitives of the model so that each dimension has its

[19]See The Economist (2018) for a discussion of the self-regulation of licensing by the legal profession in the United States. Also the American Medical Association detailing their campaigning to prevent 'scope creep' (Smith, 2025)

[20]See Cummins (1926) for an example of historical disputes involving the United Brotherhood of Carpenters and Joiners of America.

own type variable as long as these variables are co-monotonically distributed.[21] In general, allowing multiple type dimensions opens up the difficulties associated with solving multi-dimensional screening problems.[22] This can make the rational benchmark very complicated and thus prevents easy comparisons with the principal's choice of mechanism under narrow inference. However, I conjecture that versions of the results continue to hold if we restrict the principal to choosing an additive transfer function and the types are positively correlated across dimensions.

Homogeneity of Cognitive Type

The analysis of this paper is restricted to the cases where all types of agents are rational agents who fully understand the transfer function or all types of agents perform narrow inference. Analysis of the case with a mixture of both rational and boundedly rational agents is much more difficult as we lose the tractability that comes from being able to write the principal's objective and all the constraints in terms of the beliefs of the narrow agents only.[23] With mixed cognitive types we would have to account for the beliefs of the rational and narrow agents about the transfer function when solving the problem, and the relationship between them would become one of the constraints. For any action a_i on a given dimension we can write this relationship as

$$\bar{t}_i(a_i) = \sum_{a_{-i} \in A_{-i}} \frac{g(a_i, a_{-i})}{g_i(a_i)} t(a_i, a_{-i}) \tag{22}$$

The dependence of this relationship on the equilibrium distribution over actions g makes it complicated to solve for the principal's optimal mechanism under this nonlinear constraint. As such I leave studying narrow inference with mixed cognitive types to future work.

Restriction to Action Dependent Mechanisms

In the paper I restrict the principal to choose from indirect mechanisms that only depend on the agents actions and do not allow more complicated type reporting

[21] Under this transformation the action $a_i = 1$ on any dimension has a type-independent predictable utility $q \in [v_i(0), v_i(1)]$ that is distributed according to cdf $\tilde{P}_i(q) = P(v_i^{-1}(q))$.

[22] For examples, see (Rochet, 1987) or (Armstrong, 1996).

[23] The proof of Theorem 1 outlines the steps allowing this.

structures. This fits well with applications to wage and employment contracting and is easy to reconcile with narrow inference. Under narrow inference, the agents 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 agents' attention to the joint multidimensional nature of their problem and undoing the narrow inference. Thus considering more general mechanisms requires modelling of how the agents propensity to do narrow inference is affected by the structure of the mechanism.

Allowing nonbinary actions and Heterogeneous Inference Across Dimensions

The results can be extended to allow for an arbitrary finite number of actions on each dimension, at the cost of a more convoluted exposition. A more interesting extension would be to allow for different inference procedures across dimensions. For example, for each decision that they face the agent could have a different — potentially misspecified — model of how the transfer varies with a subset of variables. These models could be inconsistent, reflecting the fact the agent is modelling the world in a 'narrow' way.

A Appendix

Proof of Proposition 1

As stated, this result adapts Proposition 3.1 of Carroll (2017) to the model of this paper. We use the following two lemmas in the proof

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

$$U(s) = U(s') + \sum_{i\in N} \int_{s'}^{s} v_i'(z) \sum_{a_i\in A_i} a_i g(a_i|z) dz \tag{23}$$

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

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

We can use these rewritten ICs to show the envelope condition holds by using the Lipschitz continuity arguments in Theorems 1 and 2 of Milgrom and Segal (2002). □

Lemma A.2. *Let g^* be a 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*

$$t(a_1, ..., a_n) = \sum_{i\in N} t^i(a_i) \tag{24}$$

$$t^i(0) = \frac{1}{n} U(0), t^i(1) = -v_i(\hat{s}_i) + \frac{1}{n} U(0) \text{ for all } i \in N \tag{25}$$

Proof. The expected utility of type s behaving according to 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 $U(0)$. □

We can rewrite the principal's objective (2) using the envelope formula (23) from Lemma A.1 and the expression for the transfer function in terms of utilities in the direct mechanism.

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

Where the first line follows from expressing the transfer function in terms of expected utility and the last line follows from a standard switching of the order of integration.

Clearly it is optimal to set the expected utility of the lowest type to zero, which is consistent with participation constraint $t(0) \geq 0$. We now consider a relaxed version of the Principal's problem where we ignore the requirements of IC and that expected utilities might not be achieved given g. We show that under IVV, the mechanism that solves this relaxed problem implements a threshold strategy. By Lemma A.2 we can find a transfer function that implements the threshold strategy as IC 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.

$$\max_{g_i\in\Delta(A_i)^S}\int_0^1 [\sum_{i\in N, a_i\in A_i}(\phi_i(s)+w_i(s))a_i g_i(a_i|s)]\,p(s)\,ds \tag{26}$$

This problem can be solved pointwise by strategy $g_i(a_i|s) = 1$ if and only if $a_i \in \arg\max_{\tilde{a}_i\in A_i}(\phi_i(s)+w_i(s))\tilde{a}_i$. By IVV, $\phi_i(s)+w_i(s)$ is strictly increasing in $s \in S$, and thus there is some threshold $\hat{s}_i \in [0,1]$ such that the threshold strategy $g_i(a_i|s) = \mathbb{1}\{s \geq \hat{s}_i\}$ maximizes the objective.

Proof of Proposition 2

By Lemma 2, the beliefs inducing any threshold strategy must satisfy

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

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 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 it is defined on all action combinations, whilst preserving additivity.

Lemma A.3. *Let $\tilde{g}$ be a 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 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 that we can find $a'' \neq a'$, $\tilde{g}(a''|s'') > 0$ and $\tilde{g}(a'|s') > 0$ for some $s'' > s'$ 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 $s'' \geq \hat{s}_j > s'$ must hold, a contradiction. □

This implies that 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 $A(g) = \{a \in A : \exists s \in S \text{ such that } g(a|s) > 0\}$. 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 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 $a \in A(g)$, $\cup_{\tilde{a}_i \in A_i} \mathcal{A}_i(\tilde{a}_i) = A(g)$ 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 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 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 an $\tilde{a}(i) \in \mathcal{A}(a_i)$ 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 $\{\tilde{a}(1), ..., \tilde{a}(n)\}$. 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 $\tilde{a} \in \mathcal{A}_{\pi(l)}(a^l_{\pi(l)})$. Then we can recursively define a transfer function t from the beliefs $\bar{t}$ that render g NIC. For any $k \in \{2, ..., |A(g)|\}$

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

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 a 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 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 $(1, a_{-i}) \succ (0, \tilde{a}_{-i})$ nor $(0, \tilde{a}_{-i}) \succ (1, a_{-i})$. □

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 $a_j^1 \geq \tilde{a}_j$ 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 $t^i(a_i^1) = \frac{1}{n} t(a^1)$ 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 $a \in A \backslash A(g)$ is such that $g(a) = 0$ and thus $t(a)$ does not affect on-path beliefs $\overline{t}(a_i) = \sum_{a_{-i} \in A_{-i}} \frac{g(a_i, a_{-i})}{g_i(a_i)} t(a)$. For all $a \in A(g)$ any t' that implements the beliefs $\overline{t}$ 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 $\overline{t}$ 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 3 steps. First we prove the result for the case of narrow participation constraints. We then show how the proof is modified for the case of the joint participation constraint.

Step 1: We can write the principal's problem in a virtual-value form. From the statistical-correctness constraint in Lemma 1 and the result that any NIC strategy takes a threshold form in Lemma 2, for any dimension $i \in N$ we can write

$$\begin{aligned}\int_0^1 \sum_{a\in A} g(a|s)t(a)\,p(s)ds &= \sum_{a_i\in A_i} g_i(a_i)\overline{t}_i(a_i) = \int_0^1 \sum_{a_i\in A_i} \overline{t}_i(a_i)g_i(a_i|s)\,p(s)ds\\ &= (1-P(\hat{s}_i))\overline{t}_i(1) + P(\hat{s}_i)\overline{t}_i(0)\\ &= \overline{t}_i(0) - (1-P(\hat{s}_i))v_i(\hat{s}_i)\end{aligned}$$

For any NIC strategy g inducing thresholds $\hat{s}$. We then have for any $i, j \in N$ that

$$\begin{aligned}&(1-P(\hat{s}_i))\overline{t}_i(1) + P(\hat{s}_i)\overline{t}_i(0) = (1-P(\hat{s}_j))\overline{t}_j(1) + P(\hat{s}_j)\overline{t}_j(0)\\ \Leftrightarrow &\overline{t}_i(0) - (1-P(\hat{s}_i))v_i(\hat{s}_i) = \overline{t}_j(0) - (1-P(\hat{s}_j))v_j(\hat{s}_j)\end{aligned}$$

We can then write the principal's objective in terms of the transfer given for the action zero on one of the dimensions, the threshold in that dimension and the predictable utilities for the principal on all dimensions.

$$\begin{aligned}W(t,g) &= \int_0^1 \sum_{a\in A}[-t(a) + \sum_{j\in N} w_j(s)a_j]g(a|s)\,p(s)ds\\ &= -\int_0^1 \sum_{a\in A} t(a)g(a|s)\,p(s)ds + \sum_{j\in N}\int_0^1 w_j(s)a_j\,g_j(a|s)\,p(s)ds\\ &= -\overline{t}_i(0) + (1-P(\hat{s}_i))v_i(\hat{s}_i) + \sum_{j\in N}\int_{\hat{s}_j}^1 w_j(s)\,p(s)ds\\ &= -\overline{t}_i(0) + \int_{\hat{s}_i}^1 (v_i(s) - \frac{1-P(s)}{p(s)}v_i'(s))\,p(s)ds + \sum_{j\in N}\int_{\hat{s}_j}^1 w_j(s)\,p(s)ds\end{aligned} \tag{27}$$

Step 2: We obtain an upper bound to the full problem by applying the minimax theorem. Any strategy g that is NIC must have a threshold form. Any threshold strategy can be NIC for beliefs given by Lemma 2 and from Proposition 2 we can construct a transfer function t that implements these beliefs.

The principal wants to maximize $W(t,g)$ and must implement beliefs that satisfy the statistical correctness constraints. The Lagrangian of the problem for 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. As any NIC strategy g is a threshold strategy inducing thresholds $\hat{s}$, we write the value of the Lagrangian in terms of the vector of thresholds and expected transfers to the type choosing action 0 on any dimension $\bar{t}(0) = (\bar{t}_i(0))_{i\in N}$.

$$
\begin{aligned}
&\overline{W}(\hat{s}, \bar{t}(0), \lambda) \\
&= -\bar{t}_i(0) + \int_{\hat{s}_i}^{1} \phi_i(s)\, p(s)\, ds + \sum_{k\in N} \int_{\hat{s}_k}^{1} w_k(s)\, p(s)\, ds \\
&+ \sum_{j\in N\setminus\{i\}} \lambda_j \left[-\bar{t}_j(0) + \int_{\hat{s}_j}^{1} \phi_j(s)\, p(s)\, ds + \bar{t}_i(0) - \int_{\hat{s}_i}^{1} \phi_i(s)\, p(s)\, ds\right] \\
&= \Big(1 - \sum_{j\in N\setminus\{i\}} \lambda_j\Big)\left[-\bar{t}_i(0) + \int_{\hat{s}_i}^{1} \phi_i(s)\, p(s)\, ds\right] \\
&+ \sum_{j\in N\setminus\{i\}} \lambda_j \left[-\bar{t}_j(0) + \int_{\hat{s}_j}^{1} \phi_j(s)\, p(s)\, ds\right] + \sum_{k\in N} \int_{\hat{s}_k}^{1} w_k(s)\, p(s)\, ds
\end{aligned}
$$

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 $\beta_i = 1 - \sum_{j\in N\setminus\{i\}} \lambda_j$.

The narrow participation constraints require that $\bar{t}_i(0) \geq 0$ for all $i \in N$. Under a threshold strategy this reduces to the requirement that $\bar{t}_i(0) \geq 0$ for all $i \in N$. We can now write the principal's problem as follows

$$
\sup_{\hat{s}\in[0,1]^n, \bar{t}(0)\in\mathbb{R}^n} \; \min_{\beta\in\mathbb{R}^n, \sum_{j\in N}\beta_j = 1} \overline{W}(\hat{s}, \bar{t}(0), \beta)
$$
$$
\text{subject to } \bar{t}_j(0) \geq 0 \text{ for } j \in N
$$

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.

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

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 sup operator and means we have a saddle point solution.

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.

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

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

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

By the IVV assumption, this is decreasing and thus $\int_{x_j}^{1}(\beta_j\phi_j(P^{-1}(u))+w_j(P^{-1}(u)))du$ is concave in $x\in[0,1]^n$. We have that $\sum_{j\in N}\beta_j\bar{t}_j(0)$ is also concave in $\bar{t}(0)\in\mathbb{R}^n_{\geq 0}$. The sum of concave functions is also concave, thus for fixed β, $\overline{W}(P^{-1}(x),\bar{t}(0),\beta)$ is concave in $\bar{t}(0)$ and the quantiles x. Since $\overline{W}(P^{-1}(x),\bar{t}(0),\beta)$ is convex in β for fixed $(\bar{t}(0),x)$, we can apply the minimax theorem (Sion, 1958) and obtain that

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

where β^* is the minimizer. We can then rewrite this in terms of thresholds using $\hat{s}_i = P^{-1}(x_i)$.

Step 3: The minimax problem we have obtained is an upper bound to the principal's problem because we have restricted the Lagrange multipliers so that $\beta_j \in [0,1]$ for all $j \in N$. Note that this restriction is important for the minimax argument to hold, as Sion's minimax theorem requires that one of the domains of optimization is compact. Also note that statistical-correctness constraints are not necessarily affine in the thresholds $\hat{s}$, meaning Slater's condition for a saddle point in the Lagragian problem may not hold. Thus the use of the minimax theorem to obtain a saddle point here is important. We can also show that if a saddle point exists, then it must be such that $\beta_i \in [0,1]$ for all $i \in N$.[24]

We can show that the principal's value under the upper-bound minimax problem can be obtained with beliefs that satisfy all the constraints. Take the minimax solution $(\hat{s}^*, \beta^*)$ and define the following system of narrow beliefs for all $i \in N$.

$$\overline{t}_i(0) =$$
$$\int_{\hat{s}_i^*}^1 (v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s))\, p(s) ds - \sum_{j \in N} \beta_j^* \int_{\hat{s}_j^*}^1 (v_j(s) - \frac{1-P(s)}{p(s)} v_j'(s))\, p(s) ds$$
$$\overline{t}_i(1) = \overline{t}_i(0) - v_i(\hat{s}_i^*)$$

By definition of β^* we have that the narrow participation constraints hold: $\overline{t}_i(0) \geq 0$ for all $i \in N$. The narrow incentive constraints clearly hold, and the

[24] Suppose we obtained a saddle point of the full Lagrangian in Step 2 where for some dimension $j \in N$, we had that $\beta_j \notin [0,1]$. Then because $\sum_{i \in N} \beta_i = 1$ we have that either $\beta_j < 0$ or $1 - \beta_j = \sum_{i \in N \setminus \{j\}} \beta_i < 0$, which implies $\beta_k < 0$ for some $k \neq j$. Taking the dimension $h = j$ or $h = k$ for which $\beta_h < 0$, the principal could then set $\overline{t}_h(0)$ to any arbitrarily large positive number and obtain an arbitrarily large objective value. The objective value the principal can obtain in the primal problem is bounded.

statistical-correctness constraint holds for any $i, k \in N$

$$\begin{aligned}
&P(\hat{s}_i^*)\overline{t}_i(0) + (1 - P(\hat{s}_i^*))\overline{t}_i(1) \\
&= \overline{t}_i(0) - (1 - P(\hat{s}_i^*))v_i(\hat{s}_i^*) \\
&= \overline{t}_i(0) - \int_{\hat{s}_i^*}^1 (v_i(s) - \frac{1 - P(s)}{p(s)} v_i'(s))\, p(s)ds \\
&= \sum_{j \in N} \beta_j^* \int_{\hat{s}_j^*}^1 (v_j(s) - \frac{1 - P(s)}{p(s)} v_j'(s))\, p(s)ds \\
&= P(\hat{s}_k^*)\overline{t}_k(0) + (1 - P(\hat{s}_k^*))\overline{t}_k(1)
\end{aligned}$$

Plugging this belief system into the rewritten objective of the principal (27) obtained in step 1, these beliefs achieve the upper bound in the full problem as by Proposition 2 we can find a transfer function inducing these beliefs.

Modification of proof for joint participation constraint: We modify Step 2 so that in the upper-bound problem we no longer have the constraint $\overline{t}(0) \in \mathbb{R}^n_{\geq 0}$, but instead $\overline{t}(0) \in \{t \in \mathbb{R}^n : \sum_{i \in N} t_i = 0\} \equiv \tau^{SN}$. The modified upper-bound problem is

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

with

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

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 $t < 0$, we can choose $\overline{t}_j(0) = t$, and for all $l \in N \setminus \{j\}$ $\overline{t}_l(0) = -\frac{1}{n-1}t$. This satisfies the joint participation constraint and allows us to obtain an arbitrarily large payoff by choosing t.

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

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

Proof of Proposition 3

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

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

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 4

For every $i \in N$, let $\hat{s}_i^{rational}$ be the threshold such that $g_i(1|s) = \mathbb{1}\{s \geq \hat{s}_i^{rational}\}$ 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 with a narrow agent, as given by the solution in Theorem 1. Let β^* be the saddle-point shrinkage factors over dimensions from that problem under narrow participation constraints, and equal to $\frac{1}{n}$ for all $i \in N$ under the joint participation constraint.

First consider the case where $w_i(s) > 0$ for all $s \in S$. If $\hat{s}_i^{rational} = 1$, the result holds trivially. We must have $\phi_i(\hat{s}_i^{rational}) + w_i(\hat{s}_i^{rational}) \geq 0$ if $\hat{s}_i^{rational} \in [0, 1)$ with equality for an interior threshold. Thus $\phi_i(\hat{s}_i^{rational}) < 0$ if $\hat{s}_i^{rational} \in (0, 1)$ and therefore $0 \leq \phi_i(\hat{s}_i^{rational}) + w_i(\hat{s}_i^{rational}) < \beta_i^* \phi_i(\hat{s}_i^{rational}) + w_i(\hat{s}_i^{rational})$. If $\hat{s}_i^{rational} = 0$, we may have $\phi_i(\hat{s}_i^{rational}) \geq 0$, but then $0 < \beta_i^* \phi_i(\hat{s}_i^{rational}) + w_i(\hat{s}_i^{rational})$ also as $w_i(\hat{s}_i^{rational}) > 0$ by assumption. Suppose that for some dimension $j \in N$, we have $\hat{s}_j^{narrow} > \hat{s}_j^{rational}$. Since $\phi_i(s)$ is strictly increasing by IVV and $w_i(s)$ is also weakly increasing by assumption, we have that $\beta_i^* \phi_i(s) + w_i(s) > 0$ for all $s \in (\hat{s}_j^{rational}, \hat{s}_j^{narrow}]$, a contradiction as $\beta_i^* \phi_i(\hat{s}_j^{narrow}) + w_i(\hat{s}_j^{narrow}) \leq 0$ for an optimal $\hat{s}_j^{narrow} \in (0,1]$.

The second part of the result for when $w_i(s) < 0$ for all $s \in S$ holds via a symmetric argument.

Proof of Proposition 5

For the first case, for fixed thresholds $(\hat{s}_i)_{i \in M}$ we can take the expression for the principal's welfare in Theorem 1 and show that moving from merge M to the full dimension space N results in weakly higher welfare for the principal. We show the logic for the narrow participation constraint case, but it also applies with the joint participation constraint where weights are $\frac{1}{m}$ for all dimensions in the merged space and fall to $\frac{1}{n}$ in the full space.

$$
\begin{aligned}
&\min_{\beta \in [0,1]^n : \sum_{i \in N} \beta_i = 1} \overline{W}(\hat{s}; \beta : M) \\
&= \min_{\beta \in [0,1]^m : \sum_{i \in M} \beta_i = 1} \sum_{i \in M} \beta_i v_i(\hat{s}_i)(1 - P(\hat{s}_i)) + \sum_{i \in M} \int_{\hat{s}_i}^1 w_i(s)\, p(s) ds \\
&= \min_{\beta \in [0,1]^m : \sum_{i \in M} \beta_i = 1} \sum_{i \in M} \beta_i (1 - P(\hat{s}_i)) \sum_{j \in \mathcal{P}(i)} v_j(\hat{s}_j) + \sum_{i \in M} \int_{\hat{s}_i}^1 w_i(s)\, p(s) ds \\
&\leq \min_{\beta \in [0,1]^n : \sum_{j \in N} \beta_j = 1} \sum_{i \in M} (1 - P(\hat{s}_i)) \sum_{j \in \mathcal{P}(i)} \beta_j v_j(\hat{s}_j) + \sum_{i \in M} \int_{\hat{s}_i}^1 w_i(s)\, p(s) ds \\
&= \min_{\beta \in [0,1]^n : \sum_{j \in N} \beta_j = 1} \overline{W}(\hat{s}; \beta : N)
\end{aligned}
$$

Where the inequality holds as in this case for any $i \in M$ and $j \in \mathcal{P}(i)$, $0 \geq v_j(s) \geq v_i(s)$ for all $s \in S$.

For the second case, the principal also faces a loss under the merged dimension space from a reduction in the space of threshold strategies that can be implemented. This is because under the merged space for any given dimension $i \in M$, all the thresholds $\hat{s}_j$, $j \in \mathcal{P}(i)$ must be equalized.

Given a fixed threshold strategy in the full dimension space $(\hat{s}_j)_{j \in N}$, define a threshold strategy in the merged space such that $\hat{s}_i = \max_{j \in \mathcal{P}(i)} \hat{s}_j$. For any $j(i) = argmax_{k \in \mathcal{P}(i)} \hat{s}_k$ we can focus on solutions to the full dimension space minimax problem in Theorem 1 with $\beta^*_{j(i)} > 0$. If $\beta^*_{j(i)} = 0$ it would be optimal for the principal to implement the threshold $\hat{s}_{j(i)} = 1$ because $w_{j(i)}(s) < 0$ for all $s \in S$. If $\hat{s}_j = 1$ for any dimension, then as $v_k(s)(1-P(s)) > 0$ for all $k \in N$[25] and $s \in [0,1)$ at the solution to the minimax problem we can only have $\beta^*_h > 0$ for dimensions with $\hat{s}_h = 1$. Then the principal's payoff is zero under the split space, which gives the result as the principal can also obtain a zero payoff under the merged space.

With $\beta^*_{j(i)} > 0$ for every $i \in M$, we can show that the principal benefits from moving from the threshold strategy $(\hat{s}_j)_{j \in N}$ in the full dimension space to the threshold strategy with $\hat{s}_i = \max_{j \in \mathcal{P}(i)} \hat{s}_j$ for all $i \in M$ in the merged space. This then gives the result.

$$
\begin{aligned}
&\min_{\beta \in [0,1]^n : \sum_{i \in N} \beta_j = 1} \ \max_{(\hat{s}_j)_{j \in N} \in [0,1]^n} \overline{W}(\hat{s}; \beta : N) \\
&= \min_{k \in M} v_{j(k)}(\hat{s}^*_{j(k)})(1 - P(\hat{s}^*_{j(k)})) + \sum_{j \in N} \int_{\hat{s}^*_j}^1 w_j(s)\, p(s)\, ds \\
&\leq \min_{k \in M} \sum_{h \in \mathcal{P}(k)} v_h(\hat{s}^*_{j(k)})(1 - P(\hat{s}^*_{j(k)})) + \sum_{i \in M} \int_{\hat{s}^*_i}^1 \sum_{j \in \mathcal{P}(i)} w_j(s)\, p(s)\, ds \\
&= \min_{\beta \in [0,1]^m : \sum_{i \in M} \beta_i = 1} \sum_{i \in M} \beta_i v_i(\hat{s}_{j(i)})(1 - P(\hat{s}_{j(i)})) + \sum_{i \in M} \int_{\hat{s}^*_i}^1 w_i(s)\, p(s)\, ds \\
&\leq \min_{\beta \in [0,1]^m : \sum_{i \in M} \beta_i = 1} \ \max_{(\hat{s}_i)_{i \in M} \in [0,1]^m} \overline{W}(\hat{s}; \beta : M)
\end{aligned}
$$

The first equality holds because $\beta^*_{j(k)} > 0$ means that $j(k) = \arg\min_{i \in N} v_i(\hat{s}_{j(k)})(1-P(\hat{s}_{j(k)}))$. The second inequality holds because $v_h(s) \geq 0$ for all $s \in S$ and $\hat{s}_{j(i)} \geq \hat{s}_k$ for all $k \in \mathcal{P}(i)$, which weakly increases the principal's payoff as $w_k(s) < 0$ for all $s \in S$.

[25] As $v_k(s)$ is strictly increasing in s by IVV and $v_k(0) \geq 0$.

References

[1] Adams, W. J. and J. L. Yellen (1976). Commodity bundling and the burden of monopoly. *The quarterly journal of economics 90*(3), 475–498. 23

[2] Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica 64*(1), 51–75. 27

[3] Armstrong, M. (2016). Nonlinear pricing. *Annual Review of Economics 8*(1), 583–614. 22

[4] Bohren, J. A. and D. N. Hauser (2021). Learning with heterogeneous misspecified models: Characterization and robustness. *Econometrica 89*(6), 3025–3077. 8

[5] Brickley, J. A. and J. L. Zimmerman (2001). Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate finance 7*(4), 367–396. 25

[6] Camara, M. K. (2022). Computationally tractable choice. In *EC*, pp. 28. 8

[7] Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica 85*(2), 453–488. 6, 14, 15, 29, 47

[8] Cummins, E. (1926). Jurisdictional disputes of the carpenters' union. *The Quarterly Journal of Economics 40*(3), 463–494. 26

[9] Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of health economics 27*(6), 1436–1450. 25

[10] Eisenhuth, R. (2019). Reference-dependent mechanism design. *Economic Theory Bulletin 7*(1), 77–103. 9

[11] Eliaz, K. and R. Spiegler (2006). Contracting with diversely naive agents. *The Review of Economic Studies 73*(3), 689–714. 9

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

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

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

[15] Esponda, I. and D. Pouzo (2016). Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica 84*(3), 1093–1130. 8

[16] Eyster, E. and M. Rabin (2005). Cursed equilibrium. *Econometrica 73*(5), 1623–1672. 9

[17] Farhi, E. and X. Gabaix (2020). Optimal taxation with behavioral agents. *American Economic Review 110*(1), 298–336. 9

[18] 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. 9

[19] 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

[20] Frick, M., R. Iijima, and Y. Ishii (2020). Misinterpreting others and the fragility of social learning. *Econometrica 88*(6), 2281–2328. 8

[21] Frick, M., R. Iijima, and Y. Ishii (2022). Dispersed behavior and perceptions in assortative societies. *American Economic Review 112*(9), 3063–3105. 8

[22] Fudenberg, D., G. Lanzani, and P. Strack (2021). Limit points of endogenous misspecified learning. *Econometrica 89*(3), 1065–1098. 8

[23] Gallup (2018). The gig economy and alternative work arrangements. =https://www.gallup.com/workplace/240878/gig-economy-paper-2018.aspx. 25

[24] Graeber, T. (2023). Inattentive inference. *Journal of the European Economic Association 21*(2), 560–592. 8

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

[26] Heidhues, P. and B. Kőszegi (2010). Exploiting naivete about self-control in the credit market. *American Economic Review 100*(5), 2279–2303. 9

[27] Heidhues, P., B. Kőszegi, and P. Strack (2018). Unrealistic expectations and misguided learning. *Econometrica 86*(4), 1159–1214. 8

[28] Herweg, F., D. Müller, and P. Weinschenk (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review 100*(5), 2451–2477. 9

[29] Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory 123*(2), 81–104. 9, 46

[30] Jehiel, P. (2011). Manipulative auction design. *Theoretical economics 6*(2), 185–217. 9

[31] Jehiel, P. and K. Mierendorff (2024). Auction design with data-driven misspecifications: Inefficiency in private value auctions with correlation. *Theoretical Economics 19*(4), 1543–1579. 9

[32] Kleiner, M. M. and A. B. Krueger (2013). Analyzing the extent and influence of occupational licensing on the labor market. *Journal of Labor Economics 31*(S1), S173–S202. 26

[33] Kőszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature 52*(4), 1075–1118. 9

[34] 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. 8

[35] Lian, C. (2021). A theory of narrow thinking. *The Review of Economic Studies 88*(5), 2344–2374. 8

[36] Liebman, J. B. and R. J. Zeckhauser (2004). Schmeduling. *Working Paper*. 9

[37] Lockwood, B. B. (2020). Optimal income taxation with present bias. *American Economic Journal: Economic Policy 12*(4), 298–327. 9

[38] Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601. 29

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

[40] Mishel, L. (2018). Uber and the labor market: Uber drivers' compensation, wages, and the scale of uber and the gig economy. https://www.epi.org/publication/uber-and-the-labor-market-uber-drivers-compensation-wages-and-the-scale-of-uber-and-the-gig-economy/. 25

[41] Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research 6*(1), 58–73. 20, 61, 63

[42] O'Donoghue, T. and M. Rabin (2006). Optimal sin taxes. *Journal of Public Economics 90*(10-11), 1825–1849. 9

[43] Pearl, J. (2009). *Causality*. Cambridge university press. 2

[44] 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. 9

[45] Rabin, M. and G. Weizsäcker (2009). Narrow bracketing and dominated choices. *American Economic Review 99*(4), 1508–1543. 8

[46] Read, D., G. Loewenstein, M. Rabin, G. Keren, and D. Laibson (1999). Choice bracketing. *Journal of Risk and Uncertainty 19*(1/3), 171–197. 8

[47] Rees-Jones, A. and D. Taubinsky (2020). Measuring "schmeduling". *The Review of Economic Studies 87*(5), 2399–2438. 9

[48] Rochet, J.-C. (1987). A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of mathematical Economics 16*(2), 191–200. 27

[49] Rubinstein, A. (1993). On price recognition and computational complexity in a monopolistic model. *Journal of Political Economy 101*(3), 473–484. 9

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

[51] Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics 8*(1), 171 – 176. 36

[52] Smith, T. M. (2025). 4 ways the ama is working to stop scope creep. = https://www.ama-assn.org/practice-management/scope-practice/4-ways-ama-working-stop-scope-creep. 26

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

[54] Spiegler, R. (2020). Behavioral implications of causal misperceptions. *Annual Review of Economics 12*(1), 81–106. 18

[55] Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association 13*(1), 130–167. 9

[56] Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science 4*(3), 199–214. 8

[57] Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making 12*(3), 183–206. 8

[58] The Economist (2018). Occupational licensing blunts competition and boosts inequality. = https://www.economist.com/united-states/2018/02/17/occupational-licensing-blunts-competition-and-boosts-inequality. 26

[59] Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *science 211*(4481), 453–458. 8

[60] Vorjohann, P. (2024). Reference-dependent choice bracketing. *Available at SSRN: https://ssrn.com/abstract=4927177*. 8

A Additional Appendices

A.1 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 agents' 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 agents' selves has 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.

A.2 Further Analysis of Rational Benchmark

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 achieved given g are IC only if*

1. *Cyclical monotonicity condition: For any subset of types* $\{s_1, ..., s_{k+1}\} \subset S$ *with* $s_{k+1} = s_1$

$$\sum_{m=1}^{k} \sum_{i \in N} (v_i(s_{m+1}) - v_i(s_m)) \sum_{a_i \in A_i} g_i(a_i|s_{m+1}) a_i \geq 0 \tag{28}$$

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

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

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

This then implies the second condition. Iterating these rewritten ICs along any cycle of types $\{s_1, ..., s_{k+1}\}$ with $s_{k+1} = s_1$ gives that the cyclical monotonicity condition must hold. $\square$

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 cyclical monotonicity condition is satisfied. This is in contrast to the narrow agents model where the action has to be monotonic in type for each dimension. 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 we can have non-separability in an optimal selling mechanism with a co-monotonic type distribution. This is due to the 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 A.5, we show that when IVV does not hold this also applies in our model.

A.3 The Shape of the Principal's Optimal Mechanism

An implication of Theorem 1 is that if the predictable utilities of both the agent and the principal do not cross zero, then the principal's optimum can be implemented with a transfer function that loads transfers onto one dimension.

Proposition 6. *Assume the IVV assumption holds and that for all $s \in (0,1)$ and $i \in N$, $v_i(s) \neq 0$ and $w_i(s) \neq 0$. Take the solution to the principal's problem under narrow participation constraints given by Theorem 1, $(\hat{s}^*, \beta^*) \in [0,1]^n \times [0,1]^n$. Then the threshold strategy $\hat{s}^*$ is NIC and satisfies the narrow participation constraints under the following additive transfer function, for some $i^* \in N$*

$$t(a_1, ..., a_n) = \sum_{i \in N} t^i(a_i) \tag{29}$$

$$t^{i^*}(1) = -v_{i^*}(\hat{s}^*_{i^*}) \tag{30}$$

$$t^j(1) = t^j(0) = 0 \;\; \textit{for all} \;\; j \in N \setminus \{i^*\} \tag{31}$$

Proof. Since $v_i(s) \neq 0$ and $w_i(s) \neq 0$ for all $s \in (0,1)$ and $i \in N$ and these functions are continuous and weakly increasing by assumption, we must have $v_i(s) \geq 0$ or $v_i(s) \leq 0$ for all $s \in S$, $i \in N$ and $w_i(s) \geq 0$ or $w_i(s) \leq 0$ for all $s \in S$, $i \in N$.

Given $(\hat{s}^*, \beta^*)$, denote the following minimizing set of dimensions

$$\underline{N} \equiv \arg\min_{i \in N} v_i(\hat{s}^*_i)(1 - P(\hat{s}^*_i))$$

We can then denote the subset of the greatest dimensions in $\underline{N}$ by

$$\underline{N}^{max} \equiv \arg\max_{j \in \underline{N}} \hat{s}^*_j$$

Choose some $i^* \in \underline{N}^{max}$, and define the transfer function as in the statement of the proposition. First consider the case that $\hat{s}^*_{i^*} = \max_{i \in \underline{N}} \hat{s}^*_i = 1$. Then we have that $\arg\min_{i \in N} v_i(\hat{s}^*_i)(1 - P(\hat{s}^*_i)) = 0$, and all $j \in \underline{N}$ must be such that either $\hat{s}^*_j = 1$, as $v_j(\hat{s}^*_j) \neq 0$ by assumption. Any non-minimizing dimension $k \in N \setminus \{\underline{N}\}$ must have $\hat{s}^*_k = 0$ and $w_k(s) > 0$ for all $s \in S$. This is because either $w_k(s) > 0$ or $w_k(s) < 0$ for all $s \in S$ by assumption, meaning for any non-minimizing dimension we have $\hat{s}^*_k \in \{0,1\}$, and $\hat{s}^*_k \neq 1$ otherwise k would be in the set of minimizing dimensions $\underline{N}$. Since $\arg\min_{i \in N} v_i(\hat{s}^*_i)(1 - P(\hat{s}^*_i)) = 0$ it must be the case that $v_k(\hat{s}^*_k = 0) \geq 0$. This means that this threshold strategy is NIC under beliefs where for all $i \in N$ $\bar{t}_i(1) = \bar{t}_i(0) = 0$. These beliefs clearly also satisfy the narrow participation and statistical correctness constraints, and hold under the transfer function we have defined as there is no action where a_{i^*} is chosen with positive probability under this threshold strategy, as $\hat{s}^*_{i^*} = 1$.

Now consider the case where $\hat{s}^*_{i^*} = \max_{i \in \underline{N}} \hat{s}^*_i < 1$. We show that for each dimension $j \in N$, there are beliefs consistent with the proposed transfer function and threshold strategy $\hat{s}^*$ such that NIC and narrow participation holds for that dimension. We also show that the expected beliefs for any dimension must be equal to the same value; $P(\hat{s}^*_j)\overline{t}_j(0) + (1 - P(\hat{s}^*_j))\overline{t}_j(1) = -v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*}))$, which gives us that the statistical-correctness constraint holds. Together this means we have our result.

For any dimension $j \in N$ with an interior threshold, $\hat{s}^*_j \in (0,1)$, it must be the case that $j \in \underline{N}$ as $w_k(s) > 0$ or $w_k(s) < 0$ for all $s \in S$. Since $i^* \in \underline{N}^{max}$, $\hat{s}^*_{i^*}$ is a maximal interior threshold. This means that there is no action combination $\tilde{a} \in A$ such that $\tilde{a}_{i^*} = 1$ and $\tilde{a}_j = 0$ that has positive probability under the threshold strategy; $g(\tilde{a}) > 0$. Since only action combinations $a \in A$ where $a_{i^*} = 1$ have $t(a) \neq 0$ under the transfer function we have outlined, this means that the beliefs of the narrow agent on dimension j are such that $\overline{t}_j(0) = 0$ and

$$\overline{t}_j(1) = \sum_{a_{-j} \in A_{-j}} t(a_j, a_{-j}) \frac{g(a_j, a_{-j})}{g_j(a_j)} = -v_{i^*}(\hat{s}^*_{i^*}) \frac{1 - P(\hat{s}^*_{i^*})}{1 - P(\hat{s}^*_j)}$$

These beliefs satisfy NIC for this dimension as because $i^*, j \in \underline{N}^{max}$ we have that $\overline{t}_j(1) - \overline{t}_j(0) = -v_{i^*}(\hat{s}^*_{i^*}) \frac{1-P(\hat{s}^*_{i^*})}{1-P(\hat{s}^*_j)} = -v_j(\hat{s}^*_j)$. Narrow participation is also satisfied as

$$\overline{U}_j(s) = \mathbb{1}[s \geq \hat{s}^*_j](v_j(s) - v_j(\hat{s}^*_j)) \geq 0$$

The expected belief on this dimension over all types is then

$$P(\hat{s}^*_j)\overline{t}_j(0) + (1 - P(\hat{s}^*_j))\overline{t}_j(1) = -v_j(\hat{s}^*_j)(1 - P(\hat{s}^*_j)) = -v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*}))$$

When $j \in N$ is such that $\hat{s}^*_j = 1$, we must have that j is non-minimizing as $\max_{i \in \underline{N}} \hat{s}^*_i < 1$. Since all types choose $a_j = 0$ with probability one on this dimension, we have that $\overline{t}_j(0) = -v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*}))$. Narrow participation then holds as since $j \notin \underline{N}$, $\overline{t}_j(0) = -v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*})) \geq -v_j(\hat{s}^*_j)(1 - P(\hat{s}^*_j)) = 0$. We have that the expected belief is $\overline{t}_j(0) = -v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*}))$, and since $a_j = 1$ is off-path we can set $\overline{t}_j(1) = \overline{t}_j(0) - v_j(1)$ so that NIC holds.

Finally, when $j \in N$ has zero threshold $\hat{s}^*_j = 0$, then for the same reasons as in the interior case we have that $\overline{t}_j(1) = -v_{i^*}(\hat{s}^*_{i^*}) \frac{1-P(\hat{s}^*_{i^*})}{1-P(\hat{s}^*_j)} = -v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*}))$.

This is also equal to the expected belief for this dimension. Narrow participation holds as because i^* is a minimal dimension $i^* \in \underline{N}$ we have $v_j(0) \geq v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*}))$ and therefore.

$$\overline{U}_j(s) = (v_j(s) - v_{i^*}(\hat{s}^*_{i^*})(1 - P(\hat{s}^*_{i^*}))) \geq 0$$

Again, because $a_j = 0$ is off path we can set $\overline{t}_j(0) = \overline{t}_j(1) + v_j(0)$ to ensure NIC. □

When the predictable utilities cross zero for some dimension then this does not hold and the transfer function may have to load incentives onto multiple dimensions, as the following example demonstrates.

Example 2. Let $N = \{1, 2\}$ and the type distribution be uniform: $P(s) = s$. The predictable utilities take the following form for the principal: $v_i(s) = -r_i(1 - s)$ with $r_i > 0$ for all $i \in N$ and $r_2 > r_1$. For the principal let the predictable utilities be $w_1(s) = w$ for some constant $w \in (0, 2r_2)$ and $w_2(s) = s - (1 - \frac{w}{2r_2})$.

From Theorem 1, we can solve for principal's unique optimal mechanism which implements thresholds $\hat{s}_1 = 1 - \frac{w}{2r_1}$ and $\hat{s}_2 = 1 - \frac{w}{2r_2}$, with shrinkage coefficients $\beta_1 = 1$, $\beta_2 = 0$. From Proposition 2 the unique additive transfer function that implements this strategy is

$$\begin{aligned} &t(a_1, a_2) = t_1(a_1) + t_2(a_2) \\ &t_1(1) = \frac{w}{2}(1 - \frac{w}{2r_2}), \quad t_2(1) = \frac{w^2}{4r_1} \\ &t_1(0) = t_2(0) = 0 \end{aligned}$$

△

A.4 Effect of Narrow Inference on the Welfare of Agents

The shrinking size of transfers required to maintain given incentives under narrow inference may seem to in general hurt the agents' welfare when they receive transfer in equilibrium and benefit the agents when they pay transfer. However, the effects on the average welfare of agents across the type distribution can be shown to be ambiguous even in these two cases. This is because of how the principal implements different thresholds for taking actions under narrow inference than in the rational benchmark.

From Propositions 3 and 4, when actions are predictably costly to the agents and benefit the principal for all types then the principal is better off and implements lower thresholds under narrow inference. If thresholds stayed as they were under the rational benchmark, there is lower expected transfer to the agents under narrow inference and thus average agent welfare would decrease. However, the fact the principal reduces thresholds — which requires greater transfer to the agents — means that the average transfer the principal gives agents can end up actually being larger under narrow inference. For certain type distributions this means average agent welfare increases.

Similarly for the case when actions are predictably beneficial to the agents but cost the principal, the principal receives less expected transfer from the agents and implements greater thresholds. The lower expected transfer benefits the agents but fewer types of agent end up taking the beneficial actions. The effect on average agent welfare then depends on the specific parameterization.

For the purpose of constructing illustrative examples, we focus on the case where the costs and benefits of the actions are symmetric across dimensions. Clearly, in the rational benchmark the Principal can attain their maximum welfare by implementing a symmetric threshold strategy; $\hat{s}_i = \hat{s}$ for all $i \in N$. The following result shows the same thing is true under narrow inference. The result also establishes that in the symmetric case the principal's optimal mechanism under the narrow and joint participation constraints is identical.

Proposition 7. *Assume the IVV assumption holds and that $v_i(s) = v(s)$ and $w_i(s) = w(s)$ for all $i \in N$ and $s \in S$. Then we have that in both the rational benchmark and under narrow inference the principal wants to equalize the implemented thresholds across dimensions: $\hat{s}_i = \hat{s}_j$ for any $i, j \in N$.*

Proof. For the rational benchmark the result is clear from Proposition 1, while for narrow inference under the joint participation constraint it is clear from Theorem 1. For narrow inference under narrow participation constraints, take a solution to the saddle point problem in Theorem 1 $(\hat{s}^*(\beta^*), \beta^*)$. This must be lower that the welfare achieved by the same thresholds with equalized weights $\beta_i = \frac{1}{n}$ for all $i \in N$, since β^* is chosen to minimize the objective value. This is in turn lower that the welfare that can be attained by thresholds $\hat{s}_i^*(\frac{1}{n})$ maximizing the objective under the equalized weights.

$$\begin{aligned}\overline{W}(\hat{s}_i^*(\beta_i^*);\beta^*) &= \sum_{i\in N}\int_{\hat{s}_i^*(\beta_i^*)}^1 (\beta_i^*\phi(s)+w(s))\,p(s)\,ds\\ &\leq \sum_{i\in N}\int_{\hat{s}_i^*(\beta_i^*)}^1 (\frac{1}{n}\phi(s)+w(s))\,p(s)\,ds\\ &\leq \sum_{i\in N}\int_{\hat{s}_i^*(\frac{1}{n})}^1 (\frac{1}{n}\phi(s)+w(s))\,p(s)\,ds\end{aligned}$$

With equalized weights, clearly the maximizing thresholds must also be equalized; $\hat{s}_i^*(\frac{1}{n}) = \hat{s}(\frac{1}{n})$ for all $i \in N$. But these equalized weights and thresholds also solve the saddle-point problem of Theorem 1, since $\int_{\hat{s}^*(\frac{1}{n})}^1 \phi(s)\,p(s)\,ds$ is now the same across all dimensions. Thus the equalized thresholds also solve the saddle-point problem. □

Given symmetric threshold strategy $\hat{s}$ we can write average agent welfare under the rational benchmark as follows

$$AVU^{rational}(\hat{s}) = n\int_{\hat{s}}^1 v(s)\,p(s)\,ds - nv(\hat{s})(1-P(\hat{s})) \tag{32}$$

The average agent welfare under narrow inference is

$$AVU^{narrow}(\hat{s}) = n\int_{\hat{s}}^1 v(s)\,p(s)\,ds - v(\hat{s})(1-P(\hat{s})) \tag{33}$$

Since $AVU^{narrow}(\hat{s}) - AVU^{rational}(\hat{s}) = (n-1)v(\hat{s})(1-P(\hat{s}))$, for fixed thresholds the effect of narrow inference on average agent welfare depends on the sign of $v(\hat{s})$. Since the principal's choice of implemented thresholds changes in response to narrow inference, this does not pin whether narrow inference increases or decreases average agent welfare. The following examples demonstrate that for both cases where $v(s) \leq 0$ for all $s \in S$ and $v(s) \geq 0$ for all $s \in S$, average agent welfare can increase or decrease under narrow inference. Combined with Proposition 3, this shows that we can have the welfare of the principal and average agent welfare both increase under narrow inference, but also that they can both decrease.

Example 3. Let the type distribution take the piecewise form

$$P(s) = \mathbb{1}\{s \in [0,\underline{s})\}as + \mathbb{1}\{s \in [\underline{s},1]\}(1 - \frac{1-a\underline{s}}{1-\underline{s}})$$

where $a \in (0,1]$ and $\underline{s} \in [0,1]$. This nests the uniform distribution when $a = 1$. Suppose that $v_i(s) = v(s) = s - b$ for all $i \in N$, where $b \in \mathbb{R}$. If $b \leq 1$ we are in the the case where actions are always costly to agents while if $b \geq 0$ they are always beneficial.

We can calculate the virtual value under this parameterization

$$\phi(s) = v(s) - \frac{1-P(s)}{p(s)} v'(s) = \mathbb{1}\{s \in [0,\underline{s})\}(2s - b - \frac{1}{a}) + \mathbb{1}\{s \in [\underline{s},1]\}(2s - b - 1) \tag{34}$$

which is strictly increasing in s, with an upward jump at $\underline{s}$, and thus satisfies IVV. Let the utility from the actions to the principal be independent of the type; $w(s) = w$.

1. Agent welfare when actions cost the agents

Take the parameterization with $b = 1$. We are in the case fitting the application where the principal is a large organization employing the agents as workers. Narrow inference reduces the total wage costs to the organization, but also scales down the marginal cost of implementing that more types of worker acquire the costly skills by $\frac{1}{n}$. This can lead to the principal increasing the proportion of types taking the actions and paying more in wage costs in order to ensure this.

We can find parameters where this reduction can be particularly large, where the principal's optimal threshold is at the jump point $\underline{s}$ in the rational benchmark and far below the jump point but interior ($0 < s < \underline{s}$) in the narrow inference case. We can show that this occurs when the principal's utility of the actions is such that

$$\frac{1}{2}(\frac{1}{a}+1) - \underline{s} < w < \frac{1}{2}(\frac{1}{a}+1)$$
$$2(1-\underline{s}) < w < 2(1-\underline{s}) + \frac{1}{a} - 1$$

In this range, the principal's optimal threshold under narrow inference is $\hat{s}^* = \frac{1}{2}(1 + \frac{1}{a}) - w$. From our expressions for average agent welfare, (32) and (33), we can calculate the average agent welfare under the principal's optimal thresholds for the rational case and the narrow inference case as

$$AVU^{rational}(\underline{s}) = (1-\underline{s})(1-a\underline{s})$$
$$AVU^{narrow}(\hat{s^*}) = (1-a)(\underline{s}+w-\frac{1}{2}(\frac{1}{a}+1))$$

In general, there are parameters satisfying the inequalities above where average agent welfare is higher under narrow inference and where it is lower.[26] In particular if the type distribution puts a lot of mass on high types with low action costs then average agent welfare increases, as these types benefit from the greater transfer required to reduce the thresholds.

2. Agent welfare when actions benefit the agents

Now consider the case where $b = 0$. Here we are in the setting fitting the story where the principal is a self-employed worker selling services to different types of firm. By Propositions 3 and 4, in this case the principal is worse-off under narrow inference and if $w < 0$ implements higher thresholds than they do under the rational benchmark. Set $a = 1$, so the type distribution is uniform. For $-\frac{1}{2} < w < 0$, we have that the principal's optimal thresholds are $\hat{s}^{rational} = \frac{1-w}{2}$ under the rational benchmark and $\hat{s}^{narrow} = \frac{1-2w}{2}$ under narrow inference. The average agent welfare under the rational benchmark and narrow inference is then

$$AVU^{rational}(\hat{s}^{rational}) = (1-\hat{s}^{rational})^2 = \frac{1}{4}(1+w)^2$$
$$AVU^{narrow}(\hat{s}^{narrow}) = 1-\hat{s}^{narrow} = \frac{1}{2}+w$$

For $1-\sqrt{2} \leq w$ average agent welfare is at least as large in the rational benchmark as under narrow inference — due to the reduced marginal transfer the principal and extract from the agents and the subsequent reduction in types who can take the actions. However, for w close to zero average agent welfare is greater under narrow inference. The only types of agents who lose out under narrow inference are those who stop taking the action under the higher thresholds, $s \in (\frac{1-2w}{2}, \frac{1-w}{2})$. These types had strictly positive utility under the rational benchmark and switching to the zero actions means they have zero utility under narrow inference. The mass of this interval of types shrinks to zero as $w \uparrow 0$. In fact if $w = 0$, all types of agents

[26] For example, when $w = 5$, $a = 0.1$, $\underline{s} = 0.8$ then $AVU^{narrow}(\hat{s^*}) = 0.27 > AVU^{rational}(\underline{s}) = 0.184$. In contrast, when $w = 0.4$, $a = 1$, $\underline{s} = 0.8$ we have that $AVU^{narrow}(\hat{s^*}) = 0 < AVU^{rational}(\underline{s}) = 0.04$.

are at least as well-off under narrow inference as under the rational benchmark. This is because of the reduced amount of transfer the principal can extract from the agents given thresholds are unchanged. △

A.5 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 threshold strategy. Thus, since only 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 $v_i(s) \leq 0$ 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.

A.5.1 Example without IVV

Example 4. 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 $r_i \in (0, 1)$

$$v_i(s) = \begin{cases} (d_i - b_i)(1 - r_i) - \frac{d_i}{2}(1 - s) - \frac{d_i - b_i}{2}\frac{(1-r_i)^2}{1-s} & \text{if } s \in [0, r_i) \\ -\frac{b_i}{2}(1 - s) & \text{if } s \in [r_i, 1] \end{cases} \tag{35}$$

this has a continuous derivative equal to

$$v_i'(s) = \begin{cases} \frac{d_i}{2} - \frac{d_i - b_i}{2}(\frac{1-r_i}{1-s})^2 \text{ if } s \in [0, r_i) \\ \frac{b_i}{2} \text{ if } s \in [r_i, 1] \end{cases} \tag{36}$$

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.

$$v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s) = \begin{cases} (d_i - b_i)(1-r_i) - d_i(1-s) \text{ if } s \in [0, r_i) \\ -b_i(1-s) \text{ if } s \in [r_i, 1] \end{cases} \tag{37}$$

These virtual values can be decreasing for some interval of types depending on the parameters. Assume $w_i(s) = w_i$ for all $i \in N$ and 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 rational agents 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 than 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.

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

△

A.5.2 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 agents 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 that takes the action $a_i = 1$. 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 than their payoff from some threshold strategy under narrow inference.

Proposition 8. *Assume the principal is restricted to implementing a deterministic interval strategy. Under both the narrow and joint participation constraints*

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 agents are 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 agents are narrow.*

Proof. 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.

We prove the first case for narrow participation constraints. Since narrow participation constraints holding implies the joint participation constraint holds, this means the result also holds with a joint participation constraint.

Let g^{int} be an interval strategy with k intervals. Let $N_l \subseteq N$ be the subset of dimensions such that the agents take 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 takes the action $a_i = 1$ under g^{int}; $\underline{s}_i = \min\{s \in S : g_i^{int}(1|s) = 1\}$. Each $\underline{s}_i$ is the lower bound of one of the k intervals of g^{int}. For each $i \in N$ denote this interval by $[z_{l_i - 1}, z_{l_i})$. 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 (26).

$$\sum_{l=1}^{k} \sum_{i \in N_l} [(w_i(z_{l-1}) + v_i(z_{l-1}))(1 - P(z_{l-1})) - (w_i(z_l) + v_i(z_l))(1 - P(z_l))] \quad (38)$$

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

$$\begin{aligned}
&(w_i(z_{l-1}) + v_i(z_{l-1}))(1 - P(z_{l-1})) - (w_i(z_l) + v_i(z_l))(1 - P(z_l)) \\
&= (w_i(z_{l-1}) + v_i(z_{l-1}))(P(z_l) - P(z_{l-1})) - (v_i(z_l) - v_i(z_{l-1}))(1 - P(z_l)) \\
&- (w_i(z_l) - w_i(z_{l-1}))(1 - P(z_l)) \qquad (39)
\end{aligned}$$

and then rewrite the principal's welfare

$$\begin{aligned}
&\sum_{l=1}^{k} \sum_{i \in N_l} (w_i(z_{l-1}) + v_i(z_{l-1}))(P(z_l) - P(z_{l-1})) - \sum_{l=1}^{k} (1 - P(z_l)) \sum_{i \in N_l} (v_i(z_l) - v_i(z_{l-1})) \\
&- \sum_{l=1}^{k} (1 - P(z_l)) \sum_{i \in N_l} (w_i(z_l) - w_i(z_{l-1})) \qquad (40)
\end{aligned}$$

We can write the second term in (40) as

$$\begin{aligned}
&\sum_{l=1}^{k} (1 - P(z_l)) \sum_{j \in N_l} (v_j(z_l) - v_j(z_{l-1})) \\
&= \sum_{l=1}^{k} (P(z_l) - P(z_{l-1})) \sum_{m=1}^{l} \sum_{j \in N_m} (v_j(z_m) - v_j(z_{m-1})) \qquad (41)
\end{aligned}$$

For g^{int} to be IC, continuity of $v_i(.)$ and the cyclical monotonicity condition in Lemma A.1 require that for any intervals $l \in \{1, ..., k\}$ and $h \in \{1, ..., l\}$

$$\begin{aligned}
\sum_{m=h+1}^{l} \sum_{i \in N_m} (v_i(z_m) - v_i(z_{m-1})) &\geq \sum_{m=h}^{l} \sum_{j \in N_h} (v_j(z_m) - v_j(z_{m-1})) \qquad (42) \\
&= \sum_{j \in N_h} (v_j(z_l) - v_j(z_h)) \qquad (43)
\end{aligned}$$

In particular, this applies to the first interval that the action $a_{i^*} = 1$ is taken; $h = l_{i^*}$. Since for any $h \in \{1, ..., k\}$

$$\begin{aligned}
&\sum_{l=h}^{k} (P(z_l) - P(z_{l-1})) \sum_{m=h}^{l} \sum_{j \in N_h} (v_j(z_m) - v_j(z_{m-1})) \\
&= \sum_{l=h}^{k} (1 - P(z_l)) \sum_{j \in N_h} (v_j(z_l) - v_j(z_{l-1})) \qquad (44)
\end{aligned}$$

combining (41), (43), (44) and the fact $v_i(.)$ is increasing we have for any

$l \in \{1, \ldots, k\}$

$$\sum_{m=1}^{l} \sum_{j \in N_m} (v_j(z_m) - v_j(z_{m-1})) \geq 0$$

This then gives us

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

and thus from (40) the following upper bound on the welfare of the principal.

$$\begin{aligned} &\sum_{l=1}^{k} \sum_{i \in N_l} (w_i(z_{l-1}) + v_i(z_{l-1}))(P(z_l) - P(z_{l-1})) - \sum_{l=l_{i^*}}^{k} (1 - P(z_l)) \sum_{i \in N_{l_{i^*}}} (v_i(z_l) - v_i(z_{l-1})) \\ &- \sum_{l=1}^{k} (1 - P(z_l)) \sum_{i \in N_l} (w_i(z_l) - w_i(z_{l-1})) \end{aligned} \tag{45}$$

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

$$\begin{aligned} &\sum_{l=1}^{k} \sum_{i \in N_l} (w_i(z_{l-1}) + v_i(z_{l-1}))(P(z_l) - P(z_{l-1})) - \sum_{l=l_{i^*}}^{k} (v_{i^*}(z_l) - v_{i^*}(z_{l-1}))(1 - P(z_l)) \\ &- \sum_{l=1}^{k} (1 - P(z_l)) \sum_{i \in N_l} (w_i(z_l) - w_i(z_{l-1})) \\ &\leq \sum_{l=1}^{k} \sum_{i \in N_l} (w_i(z_{l-1})(1 - P(z_{l-1})) - w_i(z_l)(1 - P(z_l)) && (46) \\ &+ \sum_{l=l_{i^*}}^{k} [v_{i^*}(z_{l-1})(P(z_l) - P(z_{l-1})) - (v_{i^*}(z_l) - v_{i^*}(z_{l-1}))(1 - P(z_l))] && (47) \\ &= \sum_{j \in N} \int_{\underline{s}_j}^{1} w_j(s)\, p(s) ds + \int_{\underline{s}_{i^*}}^{1} (v_{i^*}(s) - \frac{1 - P(s)}{p(s)} v'_{i^*}(s))\, p(s) ds && (48) \end{aligned}$$

Under narrow inference, the principal can achieve this upper bound by implementing a 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 $\overline{t}_{i^*}(0) = 0$, $\overline{t}_{i^*}(1) =$

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

$$\bar{t}_j(0) = v_j(\underline{s}_j)(1 - P(\underline{s}_j)) - v_{i^*}(\underline{s}_{i^*})(1 - P(\underline{s}_{i^*})) \geq 0$$
$$\bar{t}_j(1) = \bar{t}_j(0) - v_j(\underline{s}_j)$$

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

A.6 Minimax 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.[27] For any dimension $i \in N$ and any $x \in [0, 1]$, define

$$\Phi_i(x) = \int_x^1 \phi_i(P^{-1}(u))du \tag{49}$$

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

$$\widehat{\overline{\Phi}}_i(x) = \sup\{z : (z, x) \in co(\Phi_i)\} \tag{50}$$

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 $\widehat{\overline{\Phi}}_i(.)$.[28] For any $\beta \in [0, 1]^n$ and dimension $i \in N$, define the threshold[29]

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

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

Theorem 2. *The principal maximizes their objective over all NIC mechanisms that satisfy the narrow participation constraints if and only if they choose a transfer*

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

[28] 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.

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

function that implements a threshold strategy $\hat{s}^*(\beta^*)$, where

$$\beta^* \in \underset{\beta\in[0,1]^n:\sum_{i\in N}\beta_i=1}{\arg\min} \sum_{i\in N} \int_{\hat{s}_i^*(\beta)}^1 (\beta_i \hat{\overline{\phi}}_i(s) + w_i(s))\, p(s)\, ds \tag{51}$$

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

$$\widehat{\overline{W}}(\hat{s}^*(\beta^*); \beta^*) = \sum_{i\in N} \int_{\hat{s}_i^*(\beta^*)}^1 (\beta_i^* \hat{\overline{\phi}}_i(s) + w_i(s))\, p(s)\, ds \tag{52}$$

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

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

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, $\int_x^1 (\hat{\overline{\phi}}_i(P^{-1}(u))du = \hat{\overline{\Phi}}_i(x) \geq \Phi_i(x) = \int_x^1 (\phi_i(P^{-1}(u))du$ for all $i \in N$, $x \in [0, 1]$. Since by Lemma 2 only threshold strategies are NIC and all 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 $\hat{\overline{\phi}}_i(.)$ 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

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

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

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 $\beta_i \widehat{\overline{\phi}}_i(\hat{s}_i^*(\beta)) + w_i(\hat{s}_i^*(\beta)) = 0$, 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

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

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

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

which then implies

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

giving us a minimax result for $\overline{W}(.)$ also. We can then apply Step 3 from Theorem 1 to show that this upper-bound objective value can be achieved by a particular transfer function and the beliefs it induces, which completes the proof. □