Online Appendix: Narrow Inference and Incentive Design

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 Example illustrating the difference in Proposition 4 under narrow participation and limited liability constraints

Let $N = \{1, 2\}$ and suppose that the agents' utility takes the linear form $v_i(s) = d_i + b_i s$ for $i \in N$. Assume that $b_i > 0$ and $d_i + b_i < 0$, which assures that v_i is strictly increasing (so the increasing virtual value assumption holds) and $v_i(s) < 0$ for all $s \in (0, 1]$. Let the type distribution be uniform on $(0, 1]$ with an atom $P(0) > 0$ at $s = 0$, meaning

$P(s) = P(0) + (1 - P(0))s$ and $p(s) = 1 - P(0)$ for $s \in (0, 1]$. Assume the direct utility to the principal is constant across types and positive: $w_i > 0$ for all $i \in N$. As $\frac{1-P(s)}{p(s)} = 1 - s$, the atom cancels from the virtual value $\phi_i(s) = v_i(s) - (1 - s)v_i'(s)$.

In the rational case, the interior thresholds must satisfy the following conditions

$$\begin{aligned}
&\phi_i(\hat{s}_i^{rat}) + w_i = d_i + b_i \hat{s}_i^{rat} - (1 - \hat{s}_i^{rat}) b_i + w_i = 0 \\
&\Leftrightarrow \hat{s}_i^{rat} = \frac{1}{2} - \frac{d_i}{2b_i} - \frac{w_i}{2b_i} \in (0, 1).
\end{aligned}$$

In the narrow case under the limited liability at zero constraint, the principal sets a common threshold across dimensions,[1] equal to the maximizer over $\hat{s}^c \in [0, 1]$ of

$$W_{LL}(\hat{s}^c) = \max_{j \in N} v_j(\hat{s}^c)(1 - \hat{s}^c) + (w_1 + w_2)(1 - \hat{s}^c),$$

where we have divided through by the common factor $1 - P(0) > 0$. If one dimension attains the maximum $\max_j v_j(\hat{s}^c)(1 - \hat{s}^c)$ for all $\hat{s}^c \in [0, 1]$, then the maximizing dimension in the expression for W_{LL} is constant and we can calculate the optimal common threshold by taking the first order condition. Dimension 1 is the maximizing dimension for all $\hat{s}^c$ if and only if

$$\begin{aligned}
&(d_1 - d_2) + (b_1 - b_2)\hat{s}^c \geq 0 \text{ for all } \hat{s}^c \in [0, 1] \\
&\text{equivalently if} \quad d_1 \geq d_2 \text{ and } d_1 - d_2 + b_1 - b_2 \geq 0,
\end{aligned}$$

Under these conditions $W_{LL}(\hat{s}^c) = (1 - \hat{s}^c)(d_1 + b_1 \hat{s}^c + w_1 + w_2)$ is strictly concave, and we can take the FOC to solve

$$W_{LL}'(\hat{s}^{c,LL}) = 0 \Leftrightarrow \hat{s}^{c,LL} = \frac{1}{2} - \frac{d_1}{2b_1} - \frac{w_1 + w_2}{2b_1}.$$

Assuming both thresholds are interior, comparing to the rational threshold of dimension 2, we have $\hat{s}^{c,LL} > \hat{s}_2^{rat}$ whenever $(d_2 + w_2)b_1 > (d_1 + w_1 + w_2)b_2$.

[1] This common threshold maximizes the value function in Theorem 1B, although this value can only be approached by threshold strategies which make the thresholds distinct but arbitrarily close to this common threshold (see Remark).

A Numerical Example: Let

$$d_1 = d_2 = -\frac{9}{4}, \quad b_1 = 2, \quad b_2 = \frac{1}{2}, \quad w_1 = \frac{1}{2}, \quad w_2 = \frac{5}{2},$$

with $P(0)$ any value in $(0,1)$. We have that all the conditions hold: $b_i > 0$; $d_1 + b_1 = -\frac{1}{4} < 0$ and $d_2 + b_2 = -\frac{7}{4} < 0$; $w_i > 0$; and $d_1 \geq d_2$, $b_1 \geq b_2$. We can calculate the rational thresholds as

$$\hat{s}_1^{rat} = \frac{1}{2} + \frac{9}{16} - \frac{1}{8} = \frac{15}{16}, \qquad \hat{s}_2^{rat} = \frac{1}{2} + \frac{9}{4} - \frac{5}{2} = \frac{1}{4}$$

While the common threshold under narrow inference with limited liability at zero is

$$\hat{s}^{c,LL} = \frac{1}{2} + \frac{9}{16} - \frac{3}{4} = \frac{5}{16}$$

Thus

$$\hat{s}^{c,LL} = \frac{5}{16} > \frac{1}{4} = \hat{s}_2^{rat}$$

so the threshold on dimension 2 increases when moving from rational to narrow inference under limited liability at zero, even though the threshold on dimension 1 falls ($\frac{5}{16} < \frac{15}{16}$). This contrasts with part 1 of Proposition 4, under which every threshold weakly falls.

A.3 Example illustrating ambiguous effect of bundling under limited liability at zero

The following example demonstrates how merging dimensions has an ambiguous effect when we only have the limited liability at zero constraint. Take the case with two dimensions and $v_i(s) = rs$, $w_i(s) = -c_i$ with $r, c_i > 0$ for $i \in \{1,2\}$. Assume that the type distribution is uniform on $(0,1]$ with an atom $P(0) > 0$ at $s = 0$, so $P(s) = P(0) + (1 - P(0))s$ and $p(s) = 1 - P(0)$ for $s \in (0,1]$.

If $c_1 + c_2 \geq 2r$ and $c_2 > r > c_1$ then under narrow inference the principal's optimal threshold when the two dimensions are merged is $\hat{s} = 1$ giving them zero payoff.[2] When the dimensions are not merged, under narrow inference with the limited liability at zero

[2] In fact in this case when the dimensions are merged the narrow inference and rational case coincide, since there is then only one dimension.

constraint the principal can set thresholds $\hat{s}_1 = \frac{1}{2} + \frac{c_1}{2r}$, $\hat{s}_2 = 1$. This gives them a positive payoff of $r(1 - P(0))(\frac{1}{2} - \frac{c_1}{2r})^2$.

Now suppose that $c_1 = c_2 = c$ and $\frac{1}{2}r < c < r$. From Proposition 4, since $w_i < 0$ the principal's optimal mechanism sets the threshold equal to 1 on one dimension, and equal to the same threshold as in the rational benchmark on the other. Without loss of generality, this gives $\hat{s}_1 = \frac{1}{2} + \frac{c}{2r}$ and $\hat{s}_2 = 1$, resulting in a payoff to the principal of $r(1-P(0))(\frac{1}{2}-\frac{c}{2r})^2$. Under merged dimensions, the common threshold $\hat{s} = \frac{1}{2}+\frac{c}{2r} \in (0,1)$ gives the principal a positive payoff $2r(1-P(0))(\frac{1}{2}-\frac{c}{2r})^2$. Thus merging can either benefit or cost the principal under the limited liability at zero constraint.

A.4 Further Analysis of Rational Benchmark

We analyse 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 $s_{k+1} = s_1$*

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

2. *The expected utility $U(s)$ is weakly 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. □

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.6, we show that when IVV does not hold this also applies in my model.

A.5 Differences with the model of Carroll (2017)

As in Proposition 3.1 of Carroll (2017), a co-monotone type distribution can be written as equivalent to single dimension of type. In my model, 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 $\tilde{P}_i(q) = P(v_i^{-1}(q))$.

In the quantile parameterization $z = P(s)$ of the co-monotone distribution in Carroll (2017), the agent is a buyer with values $v_i(s)$ for goods $i \in N$, comonotone across goods, and the transfer t is the negative of the buyer's total payment, so that $-t^i(1) = v_i(\hat{s}_i)$ is the posted price of good i. Carroll's virtual value for good i coincides with $\phi_i(s)$. Co-monotonicity gives $F_i(v_i(s)) = P(s)$, hence $f_i(v_i(s)) = \frac{p(s)}{v_i'(s)}$ and $v_i(s) - \frac{1-F_i(v_i(s))}{f_i(v_i(s))} = \phi_i(s)$. My setting differs from his in four respects: (i) the principal derives type-dependent utility $w_i(s)$ from the actions, which lies outside the renormalization Carroll uses to accommodate type-independent production costs (ii) the type distribution has an atom of inactive types at $s = 0$ (iii) the v_i are not sign-restricted, with limited liability at zero in place of buyer IR (although limited liability at zero is equivalent to $U(0) \geq 0$). (iv) Carroll uses a weaker single crossing condition on $\phi_i(\cdot)$ whereas I make the stronger assumption that $\phi_i(\cdot)$ is strictly increasing (IVV), $w_i(\cdot)$ is weakly increasing and so $\phi_i(\cdot) + w_i(\cdot)$ is strictly increasing.

A.6 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 $v_i(s) \leq 0$, $w_i > 0$ for all $i \in N$, $s \in S$ and the principal is restricted to implement a *deterministic interval strategy*. In a deterministic interval strategy, for some $k \in \mathbb{N}$ there is a partition of the type space $z_0 = 0 < z_1 < ... < z_{k-1} < z_k = 1$ where for any $s \in [z_{l-1}, z_l)$ for $l \in \{1, ..., k-1\}$ (and $[z_{k-1}, 1]$ for $l = k$) we have $g(a|s) = 1$ for some $a \in A$.

A.6.1 Example without IVV

Example OA.1. There are two dimensions $N = \{1, 2\}$. As in Examples A.2 and A.3, let the type distribution be uniform on $(0, 1]$ with an atom $P(0) > 0$ at $s = 0$, meaning $P(s) = P(0) + (1 - P(0))s$ and $p(s) = 1 - P(0)$ for $s \in (0, 1]$. Let the predictable utility take the following form on either dimension, for some d_i, b_i and $r_i \in (0, 1)$

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

this has a continuous derivative equal to

$$v_i'(s) = \begin{cases} \frac{d_i}{2} - \frac{d_i - b_i}{2}(\frac{1-r_i}{1-s})^2 & \text{if } s \in [0, r_i) \\ \frac{b_i}{2} & \text{if } s \in [r_i, 1] \end{cases} \tag{OA.3}$$

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.

$$v_i(s) - \frac{1-P(s)}{p(s)} v_i'(s) = \begin{cases} (d_i - b_i)(1-r_i) - d_i(1-s) \text{ if } s \in [0, r_i) \\ -b_i(1-s) \text{ if } s \in [r_i, 1] \end{cases} \tag{OA.4}$$

These virtual values can be decreasing for some interval of types depending on the parameters. Let $d_1 = b_1 = 0.54$, $r_1 = 0.6$, $d_2 = -0.12$, $b_2 = 1.1$, $r_2 = 0.66$, $w_1 = 0.4266$ and $w_2 = 0.32$. I show that under these parameters the principal's optimal separable mechanism is dominated by a non-separable mechanism implementing a non-threshold strategy. With the atom types obtaining $U(0) = t(0, \ldots, 0) = 0$ in both benchmarks, all welfare expressions below are scaled by the common factor $1 - P(0) > 0$, which we divide through.

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

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

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

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

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

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.

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

△

A.6.2 Welfare of the Principal without IVV

I now show that Proposition 3 also holds without the IVV assumption, if the principal in the rational benchmark is restricted to implementing 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 $a \in A$, with the inactive types $s = 0$ taking the zero action.

The proof for the first case works as follows. For a fixed interval strategy, on each dimension we take the lowest type that both takes the action and at which the surplus $v_i(s) + w_i(s)$ from the action is nonnegative. Activity below these cut-offs contributes negative surplus to the principal's objective. Activity above them is disciplined by incentive compatibility: whenever the action on some dimension is interrupted, the predictable utility the agents forego must be compensated through the dimensions that remain active. Aggregating these constraints shows that the welfare from the interval

strategy is at most the welfare from a threshold strategy at these cut-offs under which the principal pays only the information rent of a single dimension i^* — the dimension whose rent is largest. This is exactly the value the principal can attain, or approach, under narrow inference, where the perceived incentives on every dimension load onto a common pool of transfers.

Proposition OA.1. *Assume the principal is restricted to implementing a deterministic interval strategy in the rational benchmark.*

1. *When for every $i \in N$ we have $v_i(s) \leq 0$ and $w_i(s) > 0$ for all $s \in (0,1]$, 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 $s \in (0,1]$, 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 $v_i(s) \geq 0$ for all $i \in N$ and $s \in (0,1]$: by Lemma A.1 and Lemma A.3, whose proofs do not use the IVV assumption, the supremum of the principal's objective over all NIC mechanisms is $\max_{\hat{s} \in [0,1]^n} \widehat{W}^{NP}(\hat{s})$ under the narrow participation constraints and $\max_{\hat{s} \in [0,1]^n} \widehat{W}^{LL}(\hat{s})$ under limited liability at zero. Since $V_i(\hat{s}_i) \geq 0$ for every $i \in N$ and $\hat{s}_i \in [0,1]$, both values are at most

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

For any $\hat{s} \in [0,1]^n$ this is the principal's objective value from implementing the threshold strategy $\hat{s}$ in the rational benchmark under the transfer function given in Proposition 1, and any threshold strategy is a deterministic interval strategy. Thus the principal's optimal value in the rational benchmark is at least their optimal value facing narrow agents.

We prove the first case for narrow participation constraints. Since narrow participation implies limited liability at zero, any value attainable (or approachable) under narrow participation is attainable (or approachable) under limited liability at zero alone, so the result also holds in that case.

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 both takes the action $a_i = 1$ under g^{int} and at which the surplus is nonnegative; $\underline{s}_i = \min\{s \in (0,1] : g_i^{int}(1|s) = 1 \text{ and } v_i(s) + w_i(s) \geq 0\}$, with the convention $\underline{s}_i = 1$ if this set is empty. The minimum is attained when the set is nonempty: $\{s : v_i(s) + w_i(s) \geq 0\}$ is a closed interval $[\sigma_i, 1]$ by continuity and monotonicity of $v_i + w_i$, and the set of active types is a finite union of left-closed intervals. Note that if s is active on dimension i and $s < \underline{s}_i$ then $v_i(s) + w_i(s) < 0$, while $v_i(s) + w_i(s) \geq 0$ for all $s \geq \underline{s}_i$.

If the defining set is empty for every $i \in N$, then $v_i(s) + w_i(s) < 0$ at every type active on any dimension. Since $\phi_i(s) \leq v_i(s)$, the virtual-surplus form of the principal's objective (20) is then at most zero. As the principal can always guarantee zero facing narrow agents with the transfer function $t \equiv 0$, the result follows. Otherwise, each $\underline{s}_i$ with a nonempty defining set lies in one of the k intervals of g^{int}; 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 $i^* \in \arg\min_{i \in N} v_i(\underline{s}_i)(1 - P(\underline{s}_i))$, which we may take to have a nonempty defining set since $v_i(\underline{s}_i)(1 - P(\underline{s}_i)) \leq 0 = v_j(1)(1 - P(1))$ for any j with an empty set.

For any transfer function t under which g^{int} is IC and which satisfies limited liability at zero, the principal's welfare is at most the virtual-surplus expression (20). Using $\int_{z_{l-1}}^{z_l} \phi_i(s)\, p(s)\, ds = v_i(z_{l-1})(1 - P(z_{l-1})) - v_i(z_l)(1 - P(z_l))$ on each interval, this gives

$$W(t, g^{int}) \leq \sum_{l=1}^{k} \sum_{i \in N_l} \left[v_i(z_{l-1})(1 - P(z_{l-1})) - v_i(z_l)(1 - P(z_l)) + \int_{z_{l-1}}^{z_l} w_i(s)\, p(s)\, ds \right] \tag{OA.5}$$

We can rewrite the rent terms on interval $[z_{l-1}, z_l)$ as

$$\begin{aligned} & v_i(z_{l-1})(1 - P(z_{l-1})) - v_i(z_l)(1 - P(z_l)) \\ & = v_i(z_{l-1})(P(z_l) - P(z_{l-1})) - (v_i(z_l) - v_i(z_{l-1}))(1 - P(z_l)) \end{aligned} \tag{OA.6}$$

and then rewrite the bound (OA.5) as

$$\sum_{l=1}^{k}\sum_{i\in N_l}[v_i(z_{l-1})(P(z_l)-P(z_{l-1}))+\int_{z_{l-1}}^{z_l} w_i(s)\,p(s)\,ds]$$
$$-\sum_{l=1}^{k}(1-P(z_l))\sum_{i\in N_l}(v_i(z_l)-v_i(z_{l-1})) \tag{OA.7}$$

We can write the second term in (OA.7) as

$$\sum_{l=1}^{k}(1-P(z_l))\sum_{j\in N_l}(v_j(z_l)-v_j(z_{l-1}))$$
$$=\sum_{l=1}^{k}(P(z_l)-P(z_{l-1}))\sum_{m=1}^{l-1}\sum_{j\in N_m}(v_j(z_m)-v_j(z_{m-1})) \tag{OA.8}$$

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 $h \in \{1,...,l\}$

$$\sum_{m=h+1}^{l}\sum_{i\in N_m}(v_i(z_m)-v_i(z_{m-1}))\geq \sum_{j\in N_h}(v_j(z_l)-v_j(z_h))$$
$$=\sum_{m=h+1}^{l}\sum_{j\in N_h}(v_j(z_m)-v_j(z_{m-1})) \tag{OA.9}$$

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

$$\sum_{l=h}^{k}(P(z_l)-P(z_{l-1}))\sum_{m=h}^{l-1}\sum_{j\in N_h}(v_j(z_m)-v_j(z_{m-1}))$$
$$=\sum_{l=h}^{k}(1-P(z_l))\sum_{j\in N_h}(v_j(z_l)-v_j(z_{l-1})) \tag{OA.10}$$

combining (OA.8), (OA.9), (OA.10) and the fact $v_i(.)$ is strictly increasing we have for any $l \in \{1,\ldots,k\}$

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

Applying (OA.9) with $h=l_{i^*}$ term by term in (OA.8), and dropping the nonnegative

terms with $l \leq l_{i^*}$, 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 (OA.7) the following upper bound on the welfare of the principal.

$$\begin{aligned}
&\sum_{l=1}^{k} \sum_{i \in N_l} [v_i(z_{l-1})(P(z_l) - P(z_{l-1})) + \int_{z_{l-1}}^{z_l} w_i(s)\, p(s) ds] \\
&- \sum_{l=l_{i^*}}^{k} (1 - P(z_l)) \sum_{i \in N_{l_{i^*}}} (v_i(z_l) - v_i(z_{l-1}))
\end{aligned} \tag{OA.11}$$

Since $v_i(.)$ is strictly increasing, dropping the nonnegative increments of the dimensions in $N_{l_{i^*}} \setminus \{i^*\}$ bounds this above by

$$\begin{aligned}
&\sum_{l=1}^{k} \sum_{i \in N_l} [v_i(z_{l-1})(P(z_l) - P(z_{l-1})) + \int_{z_{l-1}}^{z_l} w_i(s)\, p(s) ds] \\
&- \sum_{l=l_{i^*}}^{k} (v_{i^*}(z_l) - v_{i^*}(z_{l-1}))(1 - P(z_l))
\end{aligned} \tag{OA.12}$$

Grouping the terms of this expression by dimension, it equals

$$\begin{aligned}
&\sum_{j \in N \setminus \{i^*\}} \sum_{\{l : j \in N_l\}} [v_j(z_{l-1})(P(z_l) - P(z_{l-1})) + \int_{z_{l-1}}^{z_l} w_j(s)\, p(s) ds] \\
&+ \sum_{\{l : i^* \in N_l\}} [v_{i^*}(z_{l-1})(P(z_l) - P(z_{l-1})) + \int_{z_{l-1}}^{z_l} w_{i^*}(s)\, p(s) ds] \\
&- \sum_{l=l_{i^*}}^{k} (v_{i^*}(z_l) - v_{i^*}(z_{l-1}))(1 - P(z_l))
\end{aligned}$$

We bound the two groups in turn. Fix $j \neq i^*$. Since $v_j(.)$ is strictly increasing, $v_j(z_{l-1})(P(z_l) - P(z_{l-1})) \leq \int_{z_{l-1}}^{z_l} v_j(s)\, p(s) ds$, so each term in the sum for dimension j is at most $\int_{z_{l-1}}^{z_l} (v_j(s) + w_j(s))\, p(s) ds$. Types active on dimension j below $\underline{s}_j$ have

$v_j(s) + w_j(s) < 0$, while $v_j(s) + w_j(s) \leq w_j(s)$ as $v_j(s) \leq 0$. Therefore

$$\sum_{\{l\,:\,j \in N_l\}} [v_j(z_{l-1})(P(z_l) - P(z_{l-1})) + \int_{z_{l-1}}^{z_l} w_j(s)\, p(s) ds]$$
$$\leq \sum_{\{l\,:\,j \in N_l\}} \int_{[z_{l-1}, z_l) \cap [\underline{s}_j, 1]} w_j(s)\, p(s) ds \leq \int_{\underline{s}_j}^{1} w_j(s)\, p(s) ds$$

where the last step holds as these are disjoint subsets of $[\underline{s}_j, 1]$ and $w_j(s) \geq 0$.

For the i^* terms, first drop the active terms with $l < l_{i^*}$: all types in such intervals are active on dimension i^* and below $\underline{s}_{i^*}$, so by the argument above these terms are at most $\int_{z_{l-1}}^{z_l} (v_{i^*}(s) + w_{i^*}(s))\, p(s) ds < 0$. Next, add and subtract $v_{i^*}(z_{l-1})(P(z_l) - P(z_{l-1}))$ for the intervals $l > l_{i^*}$ with $i^* \notin N_l$. On any such interval $z_{l-1} \geq z_{l_{i^*}} > \underline{s}_{i^*}$, so $-v_{i^*}(z_{l-1}) \leq w_{i^*}(z_{l-1})$ and thus, as $w_{i^*}(.)$ is weakly increasing,

$$-v_{i^*}(z_{l-1})(P(z_l) - P(z_{l-1})) \leq w_{i^*}(z_{l-1})(P(z_l) - P(z_{l-1})) \leq \int_{z_{l-1}}^{z_l} w_{i^*}(s)\, p(s) ds$$

This means the i^* terms are at most

$$\sum_{l=l_{i^*}}^{k} [v_{i^*}(z_{l-1})(P(z_l) - P(z_{l-1})) - (v_{i^*}(z_l) - v_{i^*}(z_{l-1}))(1 - P(z_l))]$$
$$+ \sum_{l=l_{i^*}}^{k} \int_{z_{l-1}}^{z_l} w_{i^*}(s)\, p(s) ds$$
$$= v_{i^*}(z_{l_{i^*}-1})(1 - P(z_{l_{i^*}-1})) + \int_{z_{l_{i^*}-1}}^{1} w_{i^*}(s)\, p(s) ds$$
$$= \int_{z_{l_{i^*}-1}}^{1} (\phi_{i^*}(s) + w_{i^*}(s))\, p(s) ds \leq \int_{\underline{s}_{i^*}}^{1} (\phi_{i^*}(s) + w_{i^*}(s))\, p(s) ds$$

where the first equality telescopes using (OA.6), the second uses $v_{i^*}(z_{l_{i^*}-1})(1 - P(z_{l_{i^*}-1})) = \int_{z_{l_{i^*}-1}}^{1} \phi_{i^*}(s)\, p(s) ds$, and the last inequality holds because $\phi_{i^*}(s) + w_{i^*}(s) \leq v_{i^*}(s) + w_{i^*}(s) < 0$ on $[z_{l_{i^*}-1}, \underline{s}_{i^*})$, types there being active on dimension i^* but below $\underline{s}_{i^*}$. Combining the bounds on the two groups, for any transfer function t under which g^{int} is IC and which satisfies limited liability at zero

$$W(t, g^{int}) \leq v_{i^*}(\underline{s}_{i^*})(1 - P(\underline{s}_{i^*})) + \sum_{j \in N} \int_{\underline{s}_j}^{1} w_j(s)\, p(s) ds = \widehat{W}^{NP}(\underline{s})$$

where the equality holds by the definition of i^* as the minimizing dimension.

The following lemma shows we can attain the supremum in Theorem 1A in the full problem. It generalizes Corollary 1 to hold without the IVV assumption. We use this Lemma in the next steps.

Lemma A.2. *Suppose $v_i(s) \leq 0$ and $w_i(s) > 0$ for all $i \in N$ and $s \in (0,1]$. Then any maximizer of $\widehat{W}^{NP}(\hat{s})$ over $\hat{s} \in [0,1]^n$ has no coincident interior thresholds at distinct utilities. Consequently the supremum of the principal's objective over all NIC mechanisms satisfying the narrow participation constraints is attained, and equals $\max_{\hat{s}\in[0,1]^n} \widehat{W}^{NP}(\hat{s})$.*

Proof. Since $v_j(s) \leq 0$ and $v_j(.)$ is strictly increasing, $V_j(\hat{s}_j) = v_j(\hat{s}_j)(1 - P(\hat{s}_j))$ is weakly increasing and continuous in $\hat{s}_j$. Let $\hat{s}^*$ be a maximizer of $\widehat{W}^{NP}$, and suppose some dimension j has $\hat{s}_j^* \in (0,1]$ with $V_j(\hat{s}_j^*) > \min_{i\in N} V_i(\hat{s}_i^*)$. By continuity there is $\epsilon > 0$ such that $V_j(\hat{s}_j^* - \epsilon) > \min_{i\in N} V_i(\hat{s}_i^*)$. Lowering the threshold to $\hat{s}_j^* - \epsilon$ then leaves $\min_{i\in N} V_i$ unchanged while strictly increasing the objective by $\int_{\hat{s}_j^*-\epsilon}^{\hat{s}_j^*} w_j(s)\, p(s) ds > 0$, contradicting maximality. Thus every dimension j with $\hat{s}_j^* > 0$ satisfies $V_j(\hat{s}_j^*) = \min_{i\in N} V_i(\hat{s}_i^*)$. If $\hat{s}_i^* = \hat{s}_j^* \in (0,1)$ for $i \neq j$, both dimensions are minimizers, so $v_i(\hat{s}_i^*)(1 - P(\hat{s}_i^*)) = v_j(\hat{s}_j^*)(1 - P(\hat{s}_j^*))$, and as $1 - P(\hat{s}_i^*) = 1 - P(\hat{s}_j^*) > 0$ we have $v_i(\hat{s}_i^*) = v_j(\hat{s}_j^*)$. Attainment then follows from Lemma A.1 and Part 2 of Lemma A.3, whose proofs do not use the IVV assumption. □

By Lemma A.1 and Lemma A.3 — whose proofs do not use the IVV assumption — the supremum of the principal's objective over all NIC mechanisms satisfying the narrow participation constraints equals $\max_{\hat{s}\in[0,1]^n} \widehat{W}^{NP}(\hat{s}) \geq \widehat{W}^{NP}(\underline{s}) \geq W(t, g^{int})$. Moreover, since $w_i(s) > 0$ for all $i \in N$ and $s \in (0,1]$, by Lemma A.2 this supremum is attained by a NIC mechanism satisfying the narrow participation constraints, without any appeal to approximation by nearby threshold strategies. The principal facing narrow agents therefore obtains at least the welfare from any IC deterministic interval strategy in the rational benchmark, which gives the result. □

A.7 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 and we are in the narrow participation constraint regime. 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 with narrow participation constraints.

Proposition OA.2. *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 under the narrow participation constraints 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. The symmetry assumption means there are no coincident interior thresholds at distinct utilities for all threshold strategies. Thus, for narrow inference under narrow participation constraints, take a solution to the saddle point problem in Theorem 1A $(\hat{s}^*(\beta^*), \beta^*)$. This must be lower than 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 1A, 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 $\hat{s}$ and $N = \{1,2\}$ we can write average agent welfare under the rational benchmark as follows

$$AVU^{rational}(\hat{s}) = 2\int_{\hat{s}}^1 v(s)\,p(s)\,ds - 2v(\hat{s})(1-P(\hat{s})) \tag{OA.13}$$

The average agent welfare under narrow inference is

$$AVU^{narrow}(\hat{s}) = 2\int_{\hat{s}}^1 v(s)\,p(s)\,ds - v(\hat{s})(1-P(\hat{s})) \tag{OA.14}$$

Since $AVU^{narrow}(\hat{s}) - AVU^{rational}(\hat{s}) = v(\hat{s})(1-P(\hat{s}))$, 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 down 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 OA.2. Let the type distribution take a piecewise form, define

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

where $\alpha \in (0, 1]$ and $\underline{s} \in [0, 1]$. This nests the uniform distribution when $\alpha = 1$. The type distribution accommodates the atom of inactive types at $s = 0$, and so is equal to $P(s) = P(0) + (1 - P(0))\tilde{P}(s)$. Again this cancels out from the virtual value calculations, and means as in Example OA.1 all welfare we obtain here are scaled by the common factor $1 - P(0) > 0$.

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 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

$$\phi(s) = v(s) - \frac{1 - P(s)}{p(s)}v'(s) = \mathbb{1}\{s \in [0, \underline{s})\}(2s - b - \frac{1}{\alpha}) + \mathbb{1}\{s \in [\underline{s}, 1]\}(2s - b - 1) \tag{OA.15}$$

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

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

In this range, the principal's optimal threshold under narrow inference is $\hat{s}^* = \frac{1}{2}(1 + \frac{1}{\alpha}) - w$. From our expressions for average agent welfare, (OA.13) and (OA.14), we can calculate the average agent welfare under the principal's optimal thresholds for the rational case and the narrow inference case as

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

In general, there are parameters satisfying the inequalities above where average agent welfare is higher under narrow inference and where it is lower.[3] 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 $b = 0$. 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 $\alpha = 1$, 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 $1 - \sqrt{2} \geq w$ average agent welfare is at least as large in the rational benchmark as under narrow inference — due to the reduced marginal transfer the principal can 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

[3] For example, when $w = 5$, $\alpha = 0.1$, $\underline{s} = 0.8$ then $AVU^{narrow}(\hat{s^*}) = 0.27 > AVU^{rational}(\underline{s}) = 0.184$. In contrast, when $w = 0.8$, $\alpha = 0.5$, $\underline{s} = 0.8$ we have that $AVU^{narrow}(\hat{s^*}) = 0.05 < AVU^{rational}(\underline{s}) = 0.12$.

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 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. $\triangle$

References

Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica 85*(2), 453–488. 5

Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory 123*(2), 81–104. 1