Bounded Rationality and Robust Mechanism Design: An Axiomatic Approach

By LUYAO ZHANG AND DAN LEVIN *

Online Appendices

I. Appendix A

The crucial part of the proof is that (i) implies (ii). First, since Axiom 1, 4, 5 implies von Neumann-Morgenstern's three axioms on lotteries, it follows directly from their theory (and the fact that F^c and Z are isomorphic) that there exists an affine function $u : Z \to R$, such that for all $p, q \in F^c : p \succsim q$ iff $u(p) \geq u(q)$. Moreover, u is cardinally unique. By Axiom 2, u is not a constant function. For any constant act $f \in F^c$, $V(f) = u(f)$, satisfying (1) for any $a(f) \in [0, 1]$. So $V(f)$ calibrates the preference on F^c.

For any $f \in F \backslash F^c$, pick constant acts f^{best}, $f^{worst} \in F^c$ that always generate the most and least preferred outcomes given f is chosen. Formally, $f^{best} \in \{p | p \succsim q, \forall q \in C(f)\}$ and $f^{worst} \in \{h | h \precsim q, \forall q \in C(f)\}$. For $f \in F^e \backslash F^c$, by the definition of F^e, $f^{best} \sim f^{worst}$ which implies $u(f^{best}) = u(f^{worst})$ and by Axiom 2, $f \sim f^{best} \sim f^{worst}$. So $V(f) = u(f^{best}) = u(f^{worst})$ satisfying (1) for any $a(f) \in [0, 1]$. Hence $V(f)$ also calibrates the preference on F^e.

Finally, for $f \in F \backslash F^e$, by the definition of F^e, $f^{worst} \prec f^{best}$. And by Axiom 3, $f^{worst} \precsim f \precsim f^{best}$.

LEMMA 1: *for $f \in F \backslash F^e$, Axiom 2-5 imply there exists a unique $\beta^* \in [0, 1]$ such that $f \sim \beta^* f^{best} + (1 - \beta^*) f^{worst}$.*

PROOF:

First since $u[\beta f^{best} + (1 - \beta)\ f^{worst}] = \beta u(f^{best}) + (1 - \beta) u(f^{worst})$, so for $0 \leq a < b \leq 1$, $b f^{best} + (1 - b)\ f^{worst} \succ a f^{best} + (1 - a)\ f^{worst}$. Then it ensures that if β^* exists, it is unique.

If $f \sim f^{best}$, then $\beta^* = 1$ works. The same way around, if $f \sim f^{worst}$, then $\beta^* = 0$ works. Otherwise, $f^{worst} \prec f \prec f^{best}$. Define

$$\beta^* = \sup\{\beta \in [0, 1] : f \succeq \beta f^{best} + (1 - \beta)\ f^{worst}\}.$$

* Zhang: Ohio State University, 410 Arps Hall, 1945 N. High St, Columbus, OH 43220 (email: sunshineluyao@gmail.com); Levin: Ohio State University, 433B Arps Hall, 1945 N. High St, Columbus, OH 43210 (email: Levin.36@osu.edu). The authors thank Ariel Rubinstein, Yaron Azrieli, Paul Healy and Ian Krajbich for useful discussions on the content of this paper and much more.

Since $\beta = 0$ is in the set, we aren't taking a sup over an empty set. By the definition of β^* if $1 \geq \beta > \beta^*$, then $f \prec \beta f^{best} + (1-\beta)\ f^{worst}$. Moreover, by the same argument to prove uniqueness above, if $0 \leq \beta < \beta^*$, then $f \succ \beta f^{best} + (1-\beta)\ f^{worst}$. To see this, note that if $0 \leq \beta < \beta^*$, then there exists β' such that $0 \leq \beta < \beta' \leq \beta^*$ and $f \succsim \beta' f^{best} + \left(1-\beta'\right)\ f^{worst}$ by the definition of β^*. And $\beta < \beta'$ implies that $f \succsim \beta' f^{best} + \left(1-\beta'\right)\ f^{worst} \succ \beta f^{best} + (1-\beta)\ f^{worst}$.

There are three possibilities to consider.

(1). Suppose $\beta^* f^{best} + (1-\beta^*) f^{worst} \succ f \succ f^{worst}$, then by Axiom 5 there exists $b \in (0,1)$ such that $b\left[\beta^* f^{best} + (1-\beta^*) f^{worst}\right] + (1-b) f^{worst} = b\beta^* f^{best} + (1 - b\beta^*) f^{worst} \succ f$. But $b\beta^* < \beta^*$, so by the previous argument $f \succ b\beta^* f^{best} + (1 - b\beta^*) f^{worst}$. Contradiction.

(2). Suppose instead that $f^{best} \succ f \succ \beta^* f^{best} + (1-\beta^*) f^{worst}$. Then by Axiom 5, there exists $a \in (0,1)$ such that $f \succ a\left[\beta^* f^{best} + (1-\beta^*) f^{worst}\right] + (1-a) f^{best} = (1 - a(1-\beta^*)) f^{best} + a(1-\beta^*) f^{worst}$. Since $(1 - a(1-\beta^*)) > \beta^*$, we have from above that $(1 - a(1-\beta^*)) f^{best} + a(1-\beta^*) f^{worst} \succ f$. Contradiction.

(3). This leaves us with the third possibility (which is what we want) namely that $f \sim \beta^* f^{best} + (1-\beta^*) f^{worst}$.

Proof of lemma 1 ends.

Follows the argument of lemma 1, then $V\left(f\right) = V[\beta^* f^{best} + (1-\beta^*)\ f^{worst}]$. Since $[\beta^* f^{best} + (1-\beta^*)\ f^{worst}] \in F^c$,

$$V[\beta^* f^{best} + \left(1-\beta^*\right)\ f^{worst}] = u[\beta^* f^{best} + \left(1-\beta^*\right)\ f^{worst}]$$

Moreover, since u is affine,

$$u[\beta^* f^{best} + \left(1-\beta^*\right)\ f^{worst}] = \beta^* u(f^{best}) + (1-\beta^*) u(f^{worst}).$$

Then, by the definition of f^{best} and f^{worst},

$$\min_{p \in C(f)} u(p) = u\left(f^{worst}\right) < u(f^{best}) = \max_{p \in C(f)} u(p).$$

So

$$u[\beta^* f^{best} + \left(1-\beta^*\right)\ f^{worst}] = \beta^* \max_{p \in C(f)} u(p) + \left(1-\beta^*\right) \min_{p \in C(f)} u(p).$$

Then

$$V(f) = \beta^* \max_{p \in C(f)} u(p) + \left(1-\beta^*\right) \min_{p \in C(f)} u(p).$$

So $\alpha(f) = \beta^*$ works and is uniquely determined.

II. Appendix B

($\Longrightarrow$) If s_i^* is an *obviously dominant strategy*, then by (2) and the *obvious monotonicity axiom*, (3) is satisfied.

($\Longleftarrow$) If (3) holds, assume by contradiction that s_i^* is not an *obviously dominant strategy*. Then there exists an information set $I \in \vartheta_i(s_i^*)$, a deviating strategy $s_i' \in S_i(I)[s_i^*(I)]^c$ such that

$$\inf_{(s_{-i},r_n)\in[I]} u_i\left(s_i^*, s_{-i}, \omega_n\right) < \sup_{(s_{-i},r_n)\in[I]} u_i\left(s_i', s_{-i}, \omega_n\right).$$

Then we can find an obvious preference represented by (1) with $\alpha(s_i^*) = 0$ and $\alpha(s_i') = 1$ such that $V(s_i^*) < V(s_i')$. So $s_i^* \prec_{[I]} s_i'$. Contradiction.

III. Appendix C

Since $u_i\left(s_i^*, s_{-i}^*, \omega_n\right) \geq \inf_{\omega_n'\in\Omega_N} u_i\left(s_i^*, s_{-i}^*, \omega_n'\right)$ and $u_i\left(s_i', s_{-i}^*, \omega_n\right) \leq \sup_{(s_{-i},\omega_n')\in[I]} u_i\left(s_i', s_{-i}, \omega_n'\right)$ for any $s_i' \in S_i(I)[s_i^*(I)]^c$ and $\omega_n \in \Omega_N$, (4) implies (5).