Appendix A  Retargetability over outcome lotteries

Suppose we are interested in $d$ outcomes. Each outcome could be the visitation of an MDP state, or a trajectory, or the receipt of a physical item. In the Pac-Man example of section 2, $d = 3$ states. The agent can induce each outcome with probability 1, so let $e_i \in \mathbb{R}^3$ be the standard basis vector with probability 1 on outcome $o$ and 0 elsewhere. Then the agent chooses among outcome lotteries $C := \{e_0, e_1, e_2\}$, which we partition into $A := \{e_0\}$ and $B := \{e_1, e_2\}$.  

**Definition A.1** (Outcome lotteries). A unit vector $x \in \mathbb{R}^d$ with non-negative entries is an *outcome lottery*.  

Many decisions are made consequentially: based on the consequences of the decision, on what outcomes are brought about by an act. For example, in a deterministic Atari game, a policy induces a trajectory. A reward function and discount rate tuple $(R, \gamma)$ assigns a return to each state trajectory $\tau = s_0, s_1, \ldots$: $G(\tau) = \sum_{i=0}^{\infty} \gamma^i R(s_i)$. The relevant outcome lottery is the discounted visit distribution over future states in an Atari game, and policies are optimal or not depending on which outcome lottery is induced by the policy.

**Definition A.2** (Optimality indicator function). Let $X, C \subseteq \mathbb{R}^d$ be finite, and let $u \in \mathbb{R}^d$. $\text{IsOptimal}(X \mid C, u)$ returns 1 if $\max_{x \in X} x^\top u \geq \max_{c \in C} c^\top u$, and 0 otherwise. 

We consider decision-making procedures which take in a targeting parameter $u$. For example, the column headers of Table 2a show the 6 permutations of the utility function $u(\bullet) := 10$, $u(\bullet) := 5$, $u(\bullet) := 0$, representable as a vector $u \in \mathbb{R}^3$.

$u$ can be permuted as follows. The outcome permutation $\phi \in S_d$ inducing an $d \times d$ permutation matrix $P_\phi$ in row representation: $(P_\phi)_{ij} = 1$ if $i = \phi(j)$ and 0 otherwise. Table 2a shows that for a given utility function, $\frac{2}{3}$ of its orbit agrees that $B$ is strictly optimal over $A$.

Orbit-level incentives occur when an inequality holds for most permuted parameter choices $u'$. Table 2a demonstrates an application of Turner et al. [2021]'s results: Optimal decision-making induces orbit-level incentives for choosing Pac-Man outcomes in $B$ over outcomes in $A$.

Furthermore, Turner et al. [2021] conjectured that “larger” $B$ will imply stronger orbit-level tendencies: If going right leads to 500 times as many options as going left, then right is better than left for at least 500 times as many reward functions for which the opposite is true. We prove this conjecture with theorem D.11 in appendix D.

However, orbit-level incentives do not require optimality. One clue is that the same results hold for anti-optimal agents, since anti-optimality/utility minimization of $u$ is equivalent to maximizing $-u$. Table 2b illustrates that the same orbit guarantees hold in this case.

**Definition A.3** (Anti-optimality indicator function). Let $X, C \subseteq \mathbb{R}^d$ be finite, and let $u \in \mathbb{R}^d$. $\text{AntiOpt}(X \mid C, u)$ returns 1 if $\min_{x \in X} x^\top u \leq \min_{c \in C} c^\top u$, and 0 otherwise.

Stepping beyond expected utility maximization/minimization, Boltzmann-rational decision-making selects outcome lotteries proportional to the exponential of their expected utility.

**Definition A.4** (Boltzmann rationality [Baker et al., 2007]). For $X \subseteq C$ and temperature $T > 0$, let

$$\text{Boltzmann}_T(X \mid C, u) := \frac{\sum_{x \in X} e^{T-1}x^\top u}{\sum_{c \in C} e^{T-1}c^\top u}$$

be the probability that some element of $X$ is Boltzmann-rational.

Lastly, orbit-level tendencies occur even under decision-making procedures which partially ignore expected utility and which “don’t optimize too hard.” Satisficing agents randomly choose an outcome lottery with expected utility exceeding some threshold. Table 2d demonstrates that satisficing induces orbit-level tendencies.

**Definition A.5** (Satisficing). Let $t \in \mathbb{R}$, let $X \subseteq C \subseteq \mathbb{R}^d$ be finite. $\text{Satisfice}_t(X, C \mid u) := \frac{|X \cap \{c \in C \mid c^\top u \geq t\}|}{|\{c \in C \mid c^\top u \geq t\}|}$ is the fraction of $X$ whose value exceeds threshold $t$. $\text{Satisfice}_t(X, C \mid u)$ evaluates to 0 the denominator equals 0.

---

3Our results on outcome lotteries hold for generic $x' \in \mathbb{R}^d$, but we find it conceptually helpful to consider the non-negative unit vector case.
To state our key results, we define several technical concepts which we informally used when reasoning about $A := \{e_{a} \}$ and $B := \{e_{a}, e_{b}\}$. For optimal, anti-optimal, Boltzmann-rational, and satisficing agents, proposition A.11 proves that these tendencies hold for all targeting parameter orbits.

### A range of decision-making functions are retargetable

In MDPs, Turner et al. [2021] consider state visitation distributions which record the total discounted time steps spent in each environment state, given that the agent follows some policy $\pi$ from an initial state $s$. These visitation distributions are one kind of outcome lottery, with $d = |S|$ the number of MDP states.

In general, we suppose the agent has an objective function $u \in \mathbb{R}^d$ which maps outcomes to real numbers. In Turner et al. [2021], $u$ was a state-based reward function (and so the outcomes were states). However, we need not restrict ourselves to the MDP setting.

To state our key results, we define several technical concepts which we informally used when reasoning about $A := \{e_{a} \}$ and $B := \{e_{a}, e_{b}\}$.

---

<table>
<thead>
<tr>
<th>Utility function $u$</th>
<th>10,5,0</th>
<th>10,0,5</th>
<th>5,10,0</th>
<th>5,0,10</th>
<th>0,10,5</th>
<th>0,5,10</th>
</tr>
</thead>
<tbody>
<tr>
<td>IsOptimal $\left( {e_{a} } \mid C, u' \right)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IsOptimal $\left( {e_{a}, e_{b} } \mid C, u' \right)$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(a) Dark gray columns indicate utility function permutations $u'$ for which $\text{IsOptimal} \left( B \mid C, u' \right) > \text{IsOptimal} \left( A \mid C, u' \right)$, while white indicates that the opposite strict inequality holds.

<table>
<thead>
<tr>
<th>Utility function $u'$</th>
<th>10,5,0</th>
<th>10,0,5</th>
<th>5,10,0</th>
<th>5,0,10</th>
<th>0,10,5</th>
<th>0,5,10</th>
</tr>
</thead>
<tbody>
<tr>
<td>AntiOpt $\left( {e_{a}, e_{b} } \mid C, u' \right)$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AntiOpt $\left( {e_{a} } \mid C, u' \right)$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(b) Utility-minimizing outcome selection probability.

<table>
<thead>
<tr>
<th>Utility function $u$</th>
<th>10,5,0</th>
<th>10,0,5</th>
<th>5,10,0</th>
<th>5,0,10</th>
<th>0,10,5</th>
<th>0,5,10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boltzmann$<em>{1} \left( {e</em>{a} } \mid C, u' \right)$</td>
<td>1</td>
<td>.993</td>
<td>1</td>
<td>.007</td>
<td>.993</td>
<td>.007</td>
</tr>
<tr>
<td>Boltzmann$<em>{1} \left( {e</em>{a}, e_{b} } \mid C, u' \right)$</td>
<td>.000</td>
<td>.007</td>
<td>.000</td>
<td>.993</td>
<td>.007</td>
<td>.993</td>
</tr>
</tbody>
</table>

(c) Boltzmann selection probabilities for $T = 1$, rounded to three significant digits.

<table>
<thead>
<tr>
<th>Utility function $u'$</th>
<th>10,5,0</th>
<th>10,0,5</th>
<th>5,10,0</th>
<th>5,0,10</th>
<th>0,10,5</th>
<th>0,5,10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfice$<em>{3} \left( {e</em>{a} } \mid C, u' \right)$</td>
<td>1</td>
<td>.5</td>
<td>1</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>Satisfice$<em>{3} \left( {e</em>{a}, e_{b} } \mid C, u' \right)$</td>
<td>0</td>
<td>.5</td>
<td>0</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
</tr>
</tbody>
</table>

(d) A satisficer uniformly randomly selects an outcome lottery with expected utility greater than or equal to the threshold $t$. Here, $t = 3$. When $\text{Satisfice}_3 \left( \{e_{a} \} \mid C, u' \right) = \text{Satisfice}_3 \left( \{e_{a}, e_{b} \} \mid C, u' \right)$, the column is colored medium gray.
Definition A.6 (Similarity of vector sets). For $\phi \in S_d$ and $X \subseteq \mathbb{R}^d$, $\phi \cdot X := \{P_\phi x \mid x \in X\}$. $X' \subseteq \mathbb{R}^{|S|}$ is similar to $X$ when $\exists \phi : \phi \cdot X' = X$. $\phi$ is an involution if $\phi = \phi^{-1}$ (it either transposes states, or fixes them). $X$ contains a copy of $X'$ when $X'$ is similar to a subset of $X$ via an involution $\phi$.

Definition A.7 (Containment of set copies). Let $n$ be a positive integer, and let $A, B \subseteq \mathbb{R}^d$. We say that $B$ contains $n$ copies of $A$ when there exist involutions $\phi_1, \ldots, \phi_n \in S_d$ such that $\forall i: \phi_i \cdot A =: B_i \subseteq B$ and $\forall j \neq i: \phi_i \cdot B_j = B_j$.

Definition A.8 (Containment of set copies). $B := \{e_1 \cdot e_1, e_2 \cdot e_2\}$ contains two copies of $A := \{e_1\}$ via $\phi_1 := \mathcal{X} \leftrightarrow \mathcal{X}$ and $\phi_2 := \mathcal{X} \leftrightarrow \mathcal{X}$.

Definition A.9 (IsOptimal). Let $\phi : S_d \rightarrow \mathbb{R}^d$ be an involution. For any $A \subseteq \mathbb{R}^d$, $\phi \cdot A$ is the pushforward distribution induced by applying the random vector $p(u) := P_\phi u$ to $\mathcal{X}$.

Definition A.10 (Orbit of a probability distribution). The orbit of $D_{any}$ under the symmetric group $S_d$ is $S_d \cdot D_{any} := \{\phi \cdot D_{any} \mid \phi \in S_d\}$.

Because $B$ contains 2 copies of $A$, there are “at least two times as many ways” for $B$ to be optimal, than for $A$ to be optimal. Similarly, $B$ is “at least two times as likely” to contain an anti-rational outcome lottery for generic utility functions. As demonstrated by Table 2, the key idea is that “larger” sets (a set $B$ containing several copies of set $A$) are more likely to be chosen under a wide range of decision-making criteria.

Proposition A.11 (Orbit incentives for different rationalities). Let $A, B \subseteq C \subseteq \mathbb{R}^d$ be finite, such that $B$ contains $n$ copies of $A$ via involutions $\phi_i$ such that $\phi_i \cdot C = C$.

1. **Rational choice** [Turner et al., 2021].
   
   IsOptimal $(B \mid C, D_{any}) \geq_{\text{most}} |D_{any}| \text{IsOptimal} (A \mid C, D_{any})$.

2. **Uniformly randomly choosing an optimal lottery.** For $X \subseteq C$, let
   
   FracOptimal $(X \mid C, u) := \left[\frac{\{\arg\max_{c \in C} c^\top u\} \cap X}{\{\arg\max_{c \in C} c^\top u\}}\right]$.

   Then FracOptimal $(B \mid C, D_{any}) \geq_{\text{most}} |D_{any}| \text{FracOptimal} (A \mid C, D_{any})$.

3. **Anti-rational choice.** AntiOpt $(B \mid C, D_{any}) \geq_{\text{most}} |D_{any}| \text{AntiOpt} (A \mid C, D_{any})$.

4. **Boltzmann rationality.**
   
   Boltzmann$_T (B \mid C, D_{any}) \geq_{\text{most}} |D_{any}| \text{Boltzmann$_T$} (A \mid C, D_{any})$.

5. **Uniformly randomly drawing $k$ outcome lotteries and choosing the best.** For $X \subseteq C$, $u \in \mathbb{R}^d$, and $k \geq 1$, let
   
   \[
   \text{best-of-k}(X, C \mid u) := \mathbb{E}_{a_1, \ldots, a_k \sim \text{unif}(C)} \left[\text{FracOptimal} (X \cap \{a_1, \ldots, a_k\} \mid \{a_1, \ldots, a_k\}, u)\right].
   \]

   Then best-of-k$(B \mid C, D_{any}) \geq_{\text{most}} |D_{any}| \text{best-of-k}(A \mid C, D_{any})$.

---

8Technically, definition A.7 implies that $A$ contains $n$ copies of $A$ holds for all $n$, via $n$ applications of the identity permutation. For our purposes, this provides greater generality, as all of the relevant results still hold. Enforcing pairwise disjointness of the $B_i$ would handle these issues, but would narrow our results to not apply e.g. when the $B_i$ share a constant vector.
6. Satisficing [Simon, 1956]. Satisfice \(B \mid C, D_{\text{any}} \geq_n \theta_{\text{most: } D_{\text{any}}}, Satisfice \(A \mid C, D_{\text{any}} \).

7. Quantilizing over outcome lotteries [Taylor, 2016]. Let \(P\) be the uniform probability distribution over \(C\). For \(X \subseteq C, u \in \mathbb{R}^d\), and \(q \in (0,1)\), let \(Q_{q,P}(X \mid C, u)\) (definition B.12) return the probability that an outcome lottery in \(X\) is drawn from the top \(q\)-quantile of \(P\), sorted by expected utility under \(u\). Then \(Q_{q,P}(B \mid C, u) \geq_{n \text{most: } R^d} Q_{q,P}(A \mid C, u)\).

One retargetable class of decision-making functions are those which only account for the expected utilities of available choices.

**Definition A.12** (EU-determined functions). Let \(\mathcal{P}(\mathbb{R}^d)\) be the power set of \(\mathbb{R}^d\), and let \(f : \prod_{i=1}^m \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}\). \(f\) is an EU-determined function if there exists a family of functions \(\{g_{\omega_1, \ldots, \omega_m}\}\) such that

\[
    f(X_1, \ldots, X_m \mid u) = g_{\lfloor X_1 \rceil, \ldots, \lfloor X_m \rceil}(\left\lfloor x_1^u \right\rfloor_{x_1 \in X_1}, \ldots, \left\lfloor x_m^u \right\rfloor_{x_m \in X_m}),
\]

where \([r_i]\) is the multiset of its elements \(r_i\).

For example, let \(X \subseteq C \subseteq \mathbb{R}^d\) be finite, and consider utility function \(u \in \mathbb{R}^d\). A Boltzmann-rational agent is more likely to select outcome lotteries with greater expected utility. Formally, \(Boltzmann_T(X \mid C, u) := \sum_{x \in X} \frac{P(x^u)}{\sum_{y \in C} e^u_y\text{ }e^u_y}\) depends only on the expected utility of outcome lotteries in \(X\), relative to the expected utility of all outcome lotteries in \(C\). Therefore, \(Boltzmann_T\) is a function of expected utilities. This is why \(Boltzmann_T\) satisfies the \(\geq_{n \text{ most: } D_{\text{any}}}\) relation.

**Theorem A.13** ( Orbit tendencies occur for EU-determined decision-making functions). Let \(A, B, C \subseteq \mathbb{R}^d\) be such that \(B\) contains \(n\) copies of \(A\) via \(\phi_i\) such that \(\phi_i \circ C = C\). Let \(h : \prod_{i=1}^n \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}\) is an EU-determined function, and let \(p(X \mid u) := h(X, C \mid u)\). Suppose that \(p\) returns a probability of selecting an element of \(X\) from \(C\). Then \(p(B \mid u) \geq_{n \text{ most: } R^d} p(A \mid u)\).

The key takeaway is that decisions which are determined by expected utility are straightforwardly retargetable. By changing the targeting parameter hyperparameter, the decision-making procedure can be flexibly retargeted to choose elements of “larger” sets (in terms of set copies via definition A.7). Less abstractly, for many agent rationalities—ways of making decisions over outcome lotteries—it is generally the case that larger sets will more often be chosen over smaller sets.

For example, consider a Pac-Man playing agent choosing which environmental state cycle it should end up in. Turner et al. [2021] show that for most reward functions, average-reward maximizing agents will tend to stay alive so that they can reach a wider range of environmental cycles. However, our results show that average-reward minimizing agents also exhibit this tendency, as do Boltzmann-rational agents who assign greater probability to higher-reward cycles. Any EU-based cycle selection method will—for most reward functions—tend to choose cycles which require Pac-Man to stay alive (at first).

**Appendix B**  Theoretical results

**Definition 3.2** (Inequalities which hold for most orbit elements). Suppose \(\Theta\) is a subset of a set acted on by \(S_d\), the symmetric group on \(d\) elements. Let \(f : \{A, B\} \times \Theta \to \mathbb{R}\) and let \(n \geq 1\). We write \(f(B \mid \theta) \geq_{\text{most: } \Theta} f(A \mid \theta)\) when, for all \(\theta \in \Theta\), the following cardinality inequality holds:

\[
    |\{\theta' \in \text{Orbit}_{\Theta}(\theta) \mid f(B \mid \theta') > f(A \mid \theta')\}| \geq n |\{\theta' \in \text{Orbit}_{\Theta}(\theta) \mid f(B \mid \theta') < f(A \mid \theta')\}|.
\]

**Remark.** In stating their equivalent of definition 3.2, Turner et al. [2021] define two functions \(f_1(\theta) := f(B \mid \theta)\) and \(f_2(\theta) := f(A \mid \theta)\) (both having type signature \(f_i : \Theta \to \mathbb{R}\)). For compatibility, proofs also use this notation.

**Lemma B.1** (Limited transitivity of \(\geq_{\text{most: } \Theta}\)). Let \(f_0, f_1, f_2, f_3 : \Theta \to \mathbb{R}\), and suppose \(\Theta\) is a subset of a set acted on by \(S_d\). Suppose that \(f_1(\theta) \geq_{\text{most: } \Theta} f_2(\theta)\) and \(\forall \theta \in \Theta : f_0(\theta) \geq f_1(\theta)\) and \(f_2(\theta) \geq f_3(\theta)\). Then \(f_0(\theta) \geq_{\text{most: } \Theta} f_3(\theta)\).
Proof. Let \( \theta \in \Theta \) and let \( \text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta) := \{ \theta' \in \text{Orbit}_{(\Theta)}(\theta) \mid f_\alpha(\theta') > f_b(\theta') \} \).

\[
\begin{align*}
|\text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta)| & \geq \left| \text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta) \right| \\
& \geq n \left| \text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta) \right| \\
& \geq n \left| \text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta) \right|.
\end{align*}
\] (5)

Therefore, eq. (5) follows. By assumption,

\[
|\text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta)| \geq n \left| \text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta) \right|.
\]

_eq. (6) follows. For all \( \theta' \in \text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta) \), our assumptions on \( f_\alpha \) and \( f_b \) ensure that

\[
f_\alpha(\theta') < f_\alpha(f_b) \leq f_b(\theta'),
\]

so

\[
\text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta) \subseteq \text{Orbit}_{(\Theta, f_\alpha \succ f_b)}(\theta).
\]

Then eq. (7) follows. By eq. (7), \( f_0(\theta) \geq \frac{n}{n+1} f_3(\theta) \).

Lemma B.2 (Order inversion for \( \geq_{\text{most}} \)). Let \( f_1, f_2 : \Theta \rightarrow \mathbb{R} \), and suppose \( \Theta \) is a subset of a set acted on by \( S_d \). Suppose that \( f_1(\theta) \geq_{\text{most}} f_2(\theta) \). Then \( -f_2(\theta) \geq_{\text{most}} -f_1(\theta) \).

Proof. By definition A.10, \( f_1(\theta) \geq_{\text{most}} f_2(\theta) \) means that

\[
\left| \left\{ \theta' \in \text{Orbit}_{(\Theta)}(\theta) \mid f_1(\theta') > f_2(\theta') \right\} \right| \geq \left| \left\{ \theta' \in \text{Orbit}_{(\Theta)}(\theta) \mid f_1(\theta') < f_2(\theta') \right\} \right|.
\] (8)

Then \( -f_2(\theta) \geq_{\text{most}} -f_1(\theta) \).

Remark. Lemma B.3 generalizes Turner et al. [2021]’s lemma B.2.

Lemma B.3 (Orbital fraction which agrees on (weak) inequality). Suppose \( f_1, f_2 : \Theta \rightarrow \mathbb{R} \) are such that \( f_1(\theta) \geq_{\text{most}} f_2(\theta) \). Then for all \( \theta \in \Theta \),

\[
\left| \left\{ \theta' \in (S_d \cdot \theta) \mid f_1(\theta') \geq f_2(\theta') \right\} \right| \geq \frac{n}{n+1}.
\]

Proof. All \( \theta' \in (S_d \cdot \theta) \cap \Theta \) such that \( f_1(\theta') = f_2(\theta') \) satisfy \( f_1(\theta') \geq f_2(\theta') \). Otherwise, consider the \( \theta' \in (S_d \cdot \theta) \cap \Theta \) such that \( f_1(\theta') \neq f_2(\theta') \). By assumption, at least \( \frac{n}{n+1} \) of these \( \theta' \) satisfy \( f_1(\theta') > f_2(\theta') \), in which case \( f_1(\theta') \geq f_2(\theta') \). Then the desired inequality follows.

B.1 General results on retableable functions

Definition B.4 (Functions which are increasing under joint permutation). Suppose that \( S_d \) acts on sets \( E_1, \ldots, E_m \), and let \( f : \prod_{i=1}^m E_i \rightarrow \mathbb{R} \). \( f(X) \) is increasing under joint permutation by \( P \subseteq S_d \) when \( \forall \phi \in P : f(X) \leq f(\phi \cdot X) \). If equality always holds, then \( f \) is invariant under joint permutation by \( P \).

Lemma B.5 (Expectations of joint-permutation-increasing functions are also joint-permutation-increasing). For \( E \) which is a subset of a set acted on by \( S_d \), let \( f : E \times \mathbb{R}^d \rightarrow \mathbb{R} \) be a bounded function which is measurable on its second argument, and let \( P \subseteq S_d \). Then if \( f \mid \mathcal{D}_m \) is increasing under joint permutation by \( P \), then \( f'(X \mid D_m) := E_{u \sim \mathcal{D}_m} [f(X \mid u)] \) is increasing under joint permutation by \( P \). If \( f \) is invariant under joint permutation by \( P \), then so is \( f' \).
Proof. Let distribution \( D_{\text{any}} \) have probability measure \( F \), and let \( \phi \cdot D_{\text{any}} \) have probability measure \( F_{\phi} \).

\[
\begin{align*}
f \left( X \mid D_{\text{any}} \right) &:= \mathbb{E}_{u \sim D_{\text{any}}} \left[ f(X \mid u) \right] \\
&= \int_{\mathbb{R}^d} f(X \mid u) \, dF(u) \\
&\leq \int_{\mathbb{R}^d} f(\phi \cdot X \mid P_\phi u) \, dF(u) \\
&= \int_{\mathbb{R}^d} f(\phi \cdot X \mid u') \, |\det P_\phi| \, dF_{\phi}(u') \\
&= \int_{\mathbb{R}^d} f(\phi \cdot X \mid u') \, dF_{\phi}(u') \\
&= f'(\phi \cdot X \mid \phi \cdot D_{\text{any}}).
\end{align*}
\]  

Equation (12) holds by assumption on \( f : f(X \mid u) \leq f(\phi \cdot X \mid P_\phi u) \). Furthermore, \( f(\phi \cdot X \mid \cdot) \) is still measurable, and so the inequality holds. Equation (13) follows by the definition of \( F_{\phi} \) (definition 6.3) and by substituting \( r' := P_\phi r \). Equation (14) follows from the fact that all permutation matrices have unitary determinant.

\[ \square \]

Lemma B.6 (Closure of orbit incentives under increasing functions). Suppose that \( S_d \) acts on sets \( E_1, \ldots, E_m \) (with \( E_1 \) being a poset), and let \( P \subseteq S_d \). Let \( f_1, \ldots, f_n : \prod_{i=1}^m E_i \to \mathbb{R} \) be increasing under joint permutation by \( P \) on input \( (X_1, \ldots, X_m) \), and suppose the \( f_i \) are order-preserving with respect to \( \preceq_{E_1} \). Let \( g : \prod_{i=1}^n \mathbb{R} \to \mathbb{R} \) be monotonically increasing on each argument. Then

\[
f (X_1, \ldots, X_m) := g \left( f_1 (X_1, \ldots, X_m), \ldots, f_n (X_1, \ldots, X_m) \right)
\]

is increasing under joint permutation by \( P \) and order-preserving with respect to set inclusion on its first argument. Furthermore, if the \( f_i \) are invariant under joint permutation by \( P \), then so is \( f \).

Proof. Let \( \phi \in P \).

\[
f (X_1, \ldots, X_m) := g \left( f_1 (X_1, \ldots, X_m), \ldots, f_n (X_1, \ldots, X_m) \right) \\
\leq g \left( f_1 (\phi \cdot X_1, \ldots, \phi \cdot X_m), \ldots, f_n (\phi \cdot X_1, \ldots, \phi \cdot X_m) \right) \\
=: f (\phi \cdot X_1, \ldots, \phi \cdot X_m).
\]

Equation (18) follows because we assumed that \( f_1 (X_1, \ldots, X_m) \leq f_1 (\phi \cdot X_1, \ldots, \phi \cdot X_m) \), and because \( g \) is monotonically increasing on each argument. If the \( f_i \) are all invariant, then eq. (18) is an equality.

Similarly, suppose \( X'_1 \preceq_{E_1} X_1 \). The \( f_i \) are order-preserving on the first argument, and \( g \) is monotonically increasing on each argument. Then \( f (X'_1, \ldots, X_m) \leq f (X_1, \ldots, X_m) \). This shows that \( f \) is order-preserving on its first argument.

\[ \square \]

Remark. \( g \) could take the convex combination of its arguments, or multiply two \( f_i \) together and add them to a third \( f_3 \).

Definition 3.5 (Multiply retargetable function). Let \( \Theta \) be a subset of a set acted on by \( S_d \), and let \( f : \{A, B\} \times \Theta \to \mathbb{R} \).

\( f \) is a \( (\Theta, A \xrightarrow{\theta} B) \)-retargetable function when, for each \( \theta \in \Theta \), we can choose permutations \( \phi_1, \ldots, \phi_n \in S_d \) which satisfy the following conditions: Consider any \( \theta^A \in \mathrm{Orbit}_{\Theta, A \xrightarrow{\theta} B} (\theta) := \{ \theta^* \in \mathrm{Orbit}_{\Theta} (\theta) \mid f(A \mid \theta^*) > f(B \mid \theta^*) \} \).

1. Retargetable via \( n \) permutations. \( \forall i = 1, \ldots, n : f \left( A \mid \phi_i \cdot \theta^A \right) < f \left( B \mid \phi_i \cdot \theta^A \right) \).

2. Parameter permutation is allowed by \( \Theta \). \( \forall i : \phi_i \cdot \theta^A \in \Theta \).

3. Permuted parameters are distinct. \( \forall i \neq j : \theta^* \in \mathrm{Orbit}_{\Theta, A \xrightarrow{\theta} B} (\theta) : \phi_i \cdot \theta^A \neq \phi_j \cdot \theta^A \).

Theorem 3.6 (Multiply retargetable functions have orbit-level tendencies).
If \( f \) is \((\Theta, A \xrightarrow{n} B)\)-retargetable, then \( f(B \mid \theta) \geq_{\text{most}} f(A \mid \theta) \).

**Proof.** Let \( \theta \in \Theta \), and let \( \phi_i \cdot \text{Orbit}_{\Theta, A \rightarrow B}(\theta) := \{ \phi_i \cdot \theta \mid \theta \in \text{Orbit}_{\Theta, A \rightarrow B}(\theta) \} \).

\[
|\text{Orbit}_{\Theta, B \rightarrow A}(\theta)| \geq \bigcup_{i=1}^{n} |\phi_i \cdot \text{Orbit}_{\Theta, A \rightarrow B}(\theta)| = \sum_{i=1}^{n} |\phi_i \cdot \text{Orbit}_{\Theta, A \rightarrow B}(\theta)| = n |\text{Orbit}_{\Theta, A \rightarrow B}(\theta)| .
\]

(20)

(21)

(22)

By item 1 and item 2, \( \phi_i \cdot \phi_j \cdot \text{Orbit}_{\Theta, A \rightarrow B}(\theta) \subseteq \phi_i \cdot \text{Orbit}_{\Theta, B \rightarrow A}(\theta) \) for all \( i \). Therefore, eq. (20) holds. Equation (21) follows by the assumption that parameters are distinct, and so therefore the cosets \( \phi_i \cdot \text{Orbit}_{\Theta, A \rightarrow B}(\theta) \) and \( \phi_j \cdot \text{Orbit}_{\Theta, A \rightarrow B}(\theta) \) are pairwise disjoint for \( i \neq j \). Equation (22) follows because each \( \phi_i \) acts injectively on orbit elements.

Letting \( f_A(\theta) := f(A \mid \theta) \) and \( f_B(\theta) := f(B \mid \theta) \), the shown inequality satisfies definition 3.2. We conclude that \( f(B \mid \theta) \geq_{\text{most}} f(A \mid \theta) \).

**Definition 3.3** (Simply-retargetable function). Let \( \Theta \) be a set acted on by \( S_d \), and let \( f : \{A, B\} \times \Theta \rightarrow \mathbb{R} \). If \( \exists \phi \in S_d : \forall \theta^A \in \Theta : f(B \mid \theta^A) < f(A \mid \theta^A) \implies f(A \mid \phi \cdot \theta^A) < f(B \mid \phi \cdot \theta^A) \), then \( f \) is a \((\Theta, A \xrightarrow{\text{simple}} B)\)-retargetable function.

**Proposition 3.4** (Simply-retargetable functions have orbit-level tendencies).

If \( f \) is \((\Theta, A \xrightarrow{\text{simple}} B)\)-retargetable, then \( f(B \mid \theta) \geq_{\text{most}} f(A \mid \theta) \).

**Proof.** Given that \( f \) is a \((\Theta, A \xrightarrow{\text{simple}} B)\)-retargetable function (definition 3.3), we want to show that \( f \) is a \((\Theta, A \xrightarrow{\text{B}} B)\)-retargetable function (definition 3.5 when \( n = 1 \)). Definition 3.5’s item 1 is true by assumption. Since \( \Theta \) is acted on by \( S_d \), \( \Theta \) is closed under permutation and so definition 3.5’s item 2 holds. When \( n = 1 \), there are no \( i \neq j \), and so definition 3.5’s item 3 is tautologically true.

Then \( f \) is a \((\Theta, A \xrightarrow{\text{B}} B)\)-retargetable function; apply lemma B.7.

**B.2 Helper results on retargetable functions**

| Targeting parameter \( \theta \) | \( f(\text{[|]} \mid \theta) \) | \( f(\text{[0]} \mid \theta) \) | \( f(\text{[1]} \cdot \text{[2]} \mid \theta) \) | \( f(\text{[3]} \cdot \text{[4]} \mid \theta) \) |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| \( \theta' := 1e_1 + 3e_2 + 2e_3 \) | 1 | 0 | 0 | 0 |
| \( \phi_1 \cdot \theta' = \phi_2 \cdot \theta'' := 3e_1 + 1e_2 + 2e_3 \) | 0 | 2 | 2 | 2 |
| \( \phi_1 \cdot \theta'' := 2e_1 + 3e_2 + 1e_3 \) | 0 | 2 | 2 | 2 |
| \( \theta'' := 2e_1 + 1e_2 + 3e_3 \) | 1 | 0 | 0 | 0 |
| \( \phi_1 \cdot \theta'' := 1e_1 + 3e_2 + 2e_3 \) | 0 | 2 | 2 | 2 |
| \( \theta^* := 3e_1 + 2e_2 + 1e_3 \) | 1 | 0 | 0 | 0 |

Table 3: We reuse the Pac-Man outcome set introduced in section 2. Let \( \phi_1 := \varnothing \leftrightarrow \varnothing \), \( \phi_2 := \varnothing \leftrightarrow \varnothing \). We tabularly define a function \( f \) which meets all requirements of lemma B.7, except for item 4: letting \( j := 2 \), \( f(B_2^* \mid \phi_1 \cdot \theta^*) = 2 > 0 = f(B_2^* \mid \theta^*) \). Although \( f(B \mid \theta) \geq_{\text{most}} f(A \mid \theta) \), it is not true that \( f(B \mid \theta^*) \geq_{\text{most}} f(A \mid \theta^*) \). Therefore, item 4 is generally required.

**Lemma B.7** (Quantitative general orbit lemma). Let \( \Theta \) be a subset of a set acted on by \( S_d \), and let \( f : E \times \Theta \rightarrow \mathbb{R} \). Consider \( A, B \in E \).

For each \( \theta \in \Theta \), choose involutions \( \phi_1, \ldots, \phi_n \in S_d \). Let \( \theta^* \in \text{Orbit}_{\Theta}(\theta) \).

1. **Retargetable under parameter permutation.** There exist \( B^*_i \in E \) such that if \( f(B \mid \theta^*) < f(A \mid \theta^*) \), then \( \forall i : f(A \mid \theta^*) \leq f(B^*_i \mid \phi_i \cdot \theta^*) \).
2. \( \Theta \) is closed under certain symmetries. \( f(B \mid \theta^*) < f(A \mid \theta^*) \implies \forall \theta_i : \phi_i \cdot \theta^* \in \Theta. \)

3. \( f \) is increasing on certain inputs. \( \forall i : f(B_i \mid \theta^*) \leq f(B \mid \theta^*). \)

4. Increasing under alternate symmetries. For \( j = 1, \ldots, n \) and \( i \neq j \), if \( f(A \mid \theta^*) < f(B \mid \theta^*) \), then \( f \left( B_j^* \mid \theta^* \right) \leq f \left( B_j^* \mid \phi_i \cdot \theta^* \right). \)

If these conditions hold for all \( \theta \in \Theta \), then

\[
f(B \mid \theta) \geq_{\text{most-} \theta} f(A \mid \theta). \tag{23}
\]

**Proof.** Let \( \theta \) and \( \theta^* \) be as described in the assumptions, and let \( i \in \{1, \ldots, n\} \).

\[
f(A \mid \phi_i \cdot \theta^*) = f(A \mid \phi_i^{-1} \cdot \theta^*) \tag{24}
\]

Equation (24) follows because \( \phi_i \) is an involution. Equation (25) and eq. (28) follow by item 1. Equation (26) and eq. (29) follow by item 3. Equation (27) holds by assumption on \( \theta^* \). Then eq. (29) shows that for any \( i \), \( f(A \mid \phi_i \cdot \theta^*) < f(B \mid \phi_i \cdot \theta^*) \), satisfying definition 3.5’s item 1.

This result’s item 2 satisfies definition 3.5’s item 2. We now just need to show definition 3.5’s item 3.

**Disjointness.** Let \( \theta', \theta'' \in \text{Orbit}_{\Theta, A \to B} (\theta) \) and let \( i \neq j \). Suppose \( \phi_i \cdot \theta' = \phi_j \cdot \theta'' \). We want to show that this leads to contradiction.

\[
f(A \mid \theta'') \leq f(B_j^* \mid \phi_j \cdot \theta'') = \underbrace{f(B_j^* \mid \phi_i^{-1} \cdot \theta')}_{\text{by our assumption of item 1}} \tag{30}
\]

Equation (30) follows by our assumption of item 1. Equation (31) holds because we assumed that \( \phi_j \cdot \theta'' = \phi_i \cdot \theta' \), and the involution ensures that \( \phi_i = \phi_i^{-1} \). Equation (32) is guaranteed by our assumption of item 4, given that \( \phi_i^{-1} \cdot \theta' = \phi_i \cdot \theta'' \in \text{Orbit}_{\Theta, B \to A} (\theta) \) by the first half of this proof. Equation (33) follows by our assumption of item 3. Equation (34) follows because we assumed that \( \theta'' \in \text{Orbit}_{\Theta, A \to B} (\theta) \).

Equation (35) through eq. (39) follow by the same reasoning, switching the roles of \( \theta' \) and \( \theta'' \), and of \( i \) and \( j \). But then we have demonstrated that a quantity is strictly less than itself, a contradiction. So for all \( \theta', \theta'' \in \text{Orbit}_{\Theta, A \to B} (\theta) \), when \( i \neq j \), \( \phi_i \cdot \theta' \neq \phi_j \cdot \theta'' \).

Therefore, we have shown definition 3.5’s item 3, and so \( f \) is a \((\Theta, A \rightarrow B)\)-retargetable function. Apply theorem 3.6 in order to conclude that eq. (23) holds. \( \square \)

**Definition B.8** (Superset-of-copy containment). Let \( A, B \subseteq \mathbb{R}^d \). \( B \) contains \( n \) superset-copies \( B_j^* \) of \( A \) when there exist involutions \( \phi_1, \ldots, \phi_n \) such that \( \phi_i \cdot A \subseteq B_i^* \subseteq B \), and whenever \( i \neq j \), \( \phi_i \cdot B_j^* = B_j^* \).
Lemma B.9 (Looser sufficient conditions for orbit-level incentives). Suppose that Θ is a subset of a set acted on by S_d and is closed under permutation by S_d. Let A, B ∈ E ⊆ P(R^d). Suppose that B contains n superset-copies B^*_i ∈ E of A via ϕ_i. Suppose that f(X | θ) is increasing under joint permutation by ϕ_1,..., ϕ_n ∈ S_d for all X ∈ E, θ ∈ Θ, and suppose that ∀i : ϕ_i : A ∈ E. Suppose that f is monotonically increasing on its first argument. Then f(B | θ) ≥_{mon.: Θ} f(A | θ).

Proof. We check the conditions of lemma B.7. Let θ ∈ Θ, and let ϑ* ∈ (S_d · θ) ∩ Θ be an orbit element.

Item 1. Holds since f(A | ϑ*) ≤ f(ϕ_i · A | ϕ_i · ϑ*), ϑ*, with the first inequality by assumption of joint increasing under permutation, and the second following from monotonicity (as ϕ_i · A ⊆ B^*_i by superset copy definition B.8).

Item 2. We have ∀θ* ∈ (S_d · θ) ∩ Θ : f(B | ϑ*) < f(A | ϑ*) → ∀i = 1,..., n : ϕ_i · θ* ∈ Θ since Θ is closed under permutation.

Item 3. Holds because we assumed that f is monotonic on its first argument.

Item 4. Holds because f is increasing under joint permutation on all of its inputs X, ϑ', and definition B.8 shows that ϕ_i · B^*_j = B^*_j when i ≠ j. Combining these two steps of reasoning, for all θ' ∈ Θ, it is true that f(B^*_j | ϑ') ≤ f(ϕ_i · B^*_j | ϕ_i · ϑ') ≤ f(B^*_j | ϕ_i · ϑ').

Then apply lemma B.7.

Lemma B.10 (Hiding an argument which is invariant under certain permutations). Let E_1, E_2, Θ be subsets of sets which are acted on by S_d. Let A ∈ E_1, C ∈ E_2. Suppose there exist ϕ_1,..., ϕ_n ∈ S_d such that ϕ_i · C = C. Suppose h : E_1 × E_2 × Θ → R satisfies ∀i : h(A, C | θ) ≤ h(ϕ_i · A, ϕ_i · C | ϕ_i · θ). For any X ∈ E_1, let f(X | θ) := h(X, C | θ). Then f(A | θ) is increasing under joint permutation by ϕ_i.

Furthermore, if h is invariant under joint permutation by ϕ_i, then so is f.

Proof.

\[ f(X | θ) := h(X, C | θ) \]

\[ \leq h(ϕ_i · X, ϕ_i · C | ϕ_i · θ) \]

\[ = h(ϕ_i · X, C | ϕ_i · θ) \]

\[ = f(ϕ_i · X | ϕ_i · θ). \]

Equation (41) holds by assumption. Equation (42) follows because we assumed ϕ_i · C = C. Then f is increasing under joint permutation by ϕ_i.

If h is invariant, then eq. (41) is an equality, and so ∀i : f(X | θ) = f(ϕ_i · X | ϕ_i · θ).  

B.2.1 EU-determined functions

Lemma B.11 and lemma B.5 together extend Turner et al. [2021]’s lemma E.17 beyond functions of max_{x ∈ X_i} to any functions of cardinalities and of expected utilities of set elements.

Definition A.12 (EU-determined functions). Let P(R^d) be the power set of R^d, and let f : \prod_{i=1}^m P(R^d) × R^d → R. f is an EU-determined function if there exists a family of functions \{g_{s_1,\ldots,s_m}\} such that

\[ f(X_1, \ldots, X_m | u) = g_{|X_1| \ldots |X_m|} \left( \left[ \begin{array}{c} x_1^T u \\ \vdots \\ x_m^T u \end{array} \right]_{x_1 \in X_1} \ldots \left[ \begin{array}{c} x_m^T u \end{array} \right]_{x_m \in X_m} \right), \]

where [r_i] is the multiset of its elements r_i.

Lemma B.11 (EU-determined functions are invariant under joint permutation). Suppose that f : \prod_{i=1}^m P(R^d) × R^d → R is an EU-determined function. Then for any ϕ ∈ S_d and X_1, ..., X_m, u, we have f(X_1, ..., X_m | u) = f(ϕ · X_1, ..., ϕ · X_m | ϕ · u).
Proof. 

\[ f(X_1, \ldots, X_m \mid u) \]
\[ = g_{X_1, \ldots, X_m \mid u} \left( \begin{bmatrix} x_1 \top u \end{bmatrix}_{x_1 \in X_1}, \ldots, \begin{bmatrix} x_m \top u \end{bmatrix}_{x_m \in X_m} \right) \] (44)
\[ = g_{\phi \cdot X_1, \ldots, \phi \cdot X_m} \left( \begin{bmatrix} x_1 \top u \end{bmatrix}_{x_1 \in X_1}, \ldots, \begin{bmatrix} x_m \top u \end{bmatrix}_{x_m \in X_m} \right) \] (45)
\[ = g_{\phi \cdot X_1, \ldots, \phi \cdot X_m} \left( \begin{bmatrix} (P_\phi x_1) \top (P_\phi u) \end{bmatrix}_{x_1 \in X_1}, \ldots, \begin{bmatrix} (P_\phi x_m) \top (P_\phi u) \end{bmatrix}_{x_m \in X_m} \right) \] (46)
\[ = f(\phi \cdot X_1, \ldots, \phi \cdot X_m \mid \phi \cdot u). \] (47)

Equation (46) holds because permutations \( \phi \) act injectively on \( \mathbb{R}^d \). Equation (47) follows because

\[ I = P_\phi^{-1}P_\phi = P_\phi^2P_\phi \] by the orthogonality of permutation matrices, and \( x \top P_\phi = (P_\phi x) \top \), so \( x \top u = x \top P_\phi u = (P_\phi x) \top (P_\phi u) \). \( \square \)

**Theorem A.13** (Orbit tendencies occur for EU-determined decision-making functions). Let \( A, B, C \subseteq \mathbb{R}^d \) be such that \( B \) contains \( n \) copies of \( A \) via \( \phi_i \) such that \( \phi_i \cdot C = C \). Let \( h : \prod_{i=1}^n P(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R} \) be an EU-determined function, and let \( p(X \mid u) := h(X, C \mid u) \). Suppose that \( p \) returns a probability of selecting an element of \( X \) from \( C \). Then \( p(B \mid u) \geq_{\text{most: } \mathbb{R}^d} p(A \mid u) \).

Proof. By assumption, there exists a family of functions \( \left\{ g_i^{1,|C|} \right\} \) such that for all \( X \subseteq \mathbb{R}^d \), \( h(X, C \mid u) = g_{X,|C|} \left( \begin{bmatrix} x \top u \end{bmatrix}_{x \in X}, \begin{bmatrix} c \top u \end{bmatrix}_{c \in C} \right) \). Therefore, lemma B.11 shows that \( h(A, C \mid u) \) is invariant under joint permutation by the \( \phi_i \). Letting \( \Theta := \mathbb{R}^d \), apply lemma B.10 to conclude that \( f(X \mid u) \) is invariant under joint permutation by the \( \phi_i \).

Since \( f \) returns a probability of selecting an element of \( X \), \( f \) obeys the monotonicity probability axiom: If \( X' \subseteq X \), then \( f(X' \mid u) \leq f(X \mid u) \). Then \( f(B \mid u) \geq_{\text{most: } \mathbb{R}^d} f(A \mid u) \) by lemma B.9. \( \square \)

**B.3 Particular results on retargetable functions**

**Definition B.12** (Quantization, closed form). Let the expected utility \( q \)-quantile threshold be

\[ M_{q,P}(C \mid u) := \inf \left\{ M \in \mathbb{R} \mid \mathbb{P}_{x \sim P}(x \top u > M) \leq q \right\} \] (49)

Let \( C_{>M_{q,P}(C \mid u)} := \{ c \in C \mid c \top u > M_{q,P}(C \mid u) \} \). \( C_{=M_{q,P}(C \mid u)} \) is defined similarly. Let \( \mathbb{I}_{L(x)} \) be the predicate function returning \( 1 \) if \( L(x) \) is true and \( 0 \) otherwise. Then for \( X \subseteq C \),

\[ Q_{q,P}(X \mid C, u) := \sum_{x \in X} \frac{P(x)}{q} \left( \mathbb{I}_{x \in C_{>M_{q,P}(C \mid u)}} \frac{\mathbb{I}_{x \in C_{=M_{q,P}(C \mid u)}}}{P(C_{=M_{q,P}(C \mid u)})} \left( q - P(C_{>M_{q,P}(C \mid u)}) \right) \right) \] (50)

where the summand is defined to be \( 0 \) if \( P(x) = 0 \) and \( x \in C_{=M_{q,P}(C \mid u)} \).

**Remark.** Unlike Taylor [2016]’s or Carey [2019]’s definitions, definition B.12 is written in closed form and requires no arbitrary tie-breaking. Instead, in the case of an expected utility tie on the quantile threshold, eq. (50) allots probability to outcomes proportional to their probability under the base distribution \( P \).

Thanks to theorem A.13, we straightforwardly prove most items of proposition A.11 by just rewriting each decision-making function as an EU-determined function. Most of the proof’s length comes from showing that the functions are measurable on \( u \), which means that the results also apply for distributions over utility functions \( D_{\text{any}} \in \mathcal{D}_{\text{any}} \).

**Proposition A.11** (Orbit incentives for different rationalities). Let \( A, B \subseteq C \subseteq \mathbb{R}^d \) be finite, such that \( B \) contains \( n \) copies of \( A \) via involutions \( \phi_i \) such that \( \phi_i \cdot C = C \).
1. **Rational choice** [Turner et al., 2021].

\[
\text{IsOptimal} (B \mid C, D_{\text{any}}) \geq_{\text{most} : D_{\text{any}}} \text{IsOptimal} (A \mid C, D_{\text{any}})
\]

2. **Uniformly randomly choosing an optimal lottery.** For \( X \subseteq C \), let

\[
\text{FracOptimal} (X \mid C, u) := \frac{\{ \arg \max_{c \in C} c^\top u \} \cap X}{\{ \arg \max_{c \in C} c^\top u \}}.
\]

Then \( \text{FracOptimal} (B \mid C, D_{\text{any}}) \geq_{\text{most} : D_{\text{any}}} \text{FracOptimal} (A \mid C, D_{\text{any}}) \).

3. **Anti-rational choice.** AntiOpt \( (B \mid C, D_{\text{any}}) \geq_{\text{most} : D_{\text{any}}} \text{AntiOpt} (A \mid C, D_{\text{any}}) \).

4. **Boltzmann rationality.**

\[
\text{Boltzmann}_T (B \mid C, D_{\text{any}}) \geq_{\text{most} : D_{\text{any}}} \text{Boltzmann}_T (A \mid C, D_{\text{any}})
\]

5. **Uniformly randomly drawing \( k \) outcome lotteries and choosing the best.** For \( X \subseteq C \), \( u \in \mathbb{R}^d \), and \( k \geq 1 \), let

\[
\text{best-of-}k(X, C \mid u) := \mathbb{E}_{a_1, \ldots, a_k \sim \text{unif}(C)} \left[ \text{FracOptimal} (X \cap \{a_1, \ldots, a_k\} \mid \{a_1, \ldots, a_k\}, u) \right].
\]

Then \( \text{best-of-}k(B \mid C, D_{\text{any}}) \geq_{\text{most} : D_{\text{any}}} \text{best-of-}k(A \mid C, D_{\text{any}}) \).

6. **Satisficing** [Simon, 1956]. Satisfice \( (B \mid C, D_{\text{any}}) \geq_{\text{most} : D_{\text{any}}} \text{Satisfice} (A \mid C, D_{\text{any}}) \).

7. **Quantizing over outcome lotteries** [Taylor, 2016]. Let \( P \) be the uniform probability distribution over \( C \). For \( X \subseteq C \), \( u \in \mathbb{R}^d \), and \( q \in (0, 1] \), let \( Q_{q,P}(X \mid C, u) \) (definition B.12) return the probability that an outcome lottery in \( X \) is drawn from the top \( q \)-quantile of \( P \), sorted by expected utility under \( u \). Then \( Q_{q,P}(B \mid C, u) \geq_{\text{most} : \mathbb{R}^d} Q_{q,P}(A \mid C, u) \).

**Proof.** **Item 1.** Consider

\[
h(X, C \mid u) := \mathbb{I}_{\exists x \in X : \forall e \in C : x^\top u \geq e^\top u}
\]

\[
= \min \left( 1, \sum_{x \in X} \prod_{e \in C} \mathbb{I}_{(x-e)^\top u \geq 0} \right)
\]

Since halfspaces are measurable, each indicator function is measurable on \( u \). The finite sum of the finite product of measurable functions is also measurable. Since \( \min \) is continuous (and therefore measurable), \( h(X, C \mid u) \) is measurable on \( u \).

Furthermore, \( h \) is an EU-determined function:

\[
h(X, C \mid u) = g \left( \begin{bmatrix} V_X \vline x^\top u \vline x \in X \vline C \end{bmatrix} \begin{bmatrix} V_C \vline c^\top u \vline c \in C \end{bmatrix} \right)
\]

\[
:= \mathbb{I}_{\exists v_x \in V_X : \forall v_x \in V_C : v_x \geq v_x}
\]

Then by lemma B.11, \( h \) is invariant to joint permutation by the \( \phi_t \). Since \( \phi_t \cdot C = C \), lemma B.10 shows that \( h'(X \mid u) := h(X, C \mid u) \) is also invariant under joint permutation by the \( \phi_t \). Since \( h \) is a measurable function of \( u \), so is \( h' \). Then since \( h' \) is bounded, lemma B.5 shows that \( f(X \mid D_{\text{any}}) := \mathbb{E}_{u \sim D_{\text{any}}} \left[ h'(X \mid u) \right] \) is invariant under joint permutation by \( \phi_t \).

Furthermore, if \( X' \subseteq X \), \( f(X' \mid D_{\text{any}}) \leq f(X \mid D_{\text{any}}) \) by the monotonicity of probability. Then by lemma B.9,

\[
f(B \mid D_{\text{any}}) := \text{IsOptimal} (B \mid C, D_{\text{any}}) \geq_{\text{most} : D_{\text{any}}} \text{IsOptimal} (A \mid C, D_{\text{any}}) =: f(A \mid D_{\text{any}})
\]
We now show that \( \text{FracOptimal} \) with the Boltzmann function is also invariant to joint permutation for all permutations since \( \text{FracOptimal} \) is also invariant to joint permutation by \( \phi_i \).

Let \( \text{FracOptimal} \) be the Boltzmann function:

\[
\text{FracOptimal}(X \mid C, u) = \frac{\{\arg\max_{e \in C} e^T u\} \cap X}{\{\arg\max_{e \in C} e^T u\}}
\]

with the \( \{ \cdot \} \) denoting a multiset which allows and counts duplicates. Then by lemma B.11, \( \text{FracOptimal}(X \mid C, u) \) is invariant to joint permutation by \( \phi_i \).

We now show that \( \text{FracOptimal}(X \mid C, u) \) is a measurable function of \( u \).

\[
\text{FracOptimal}(X \mid C, u) = \frac{\{\arg\max_{e \in C} e^T u\} \cap X}{\{\arg\max_{e \in C} e^T u\}}
\]

Equation (59) holds because \( x \) belongs to the arg max iff \( \forall c \in C : \langle x \rangle^T u \geq c^T u \). Furthermore, this condition is met iff \( u \) belongs to the intersection of finitely many closed halfspaces; therefore, \( \left\{ u \in \mathbb{R}^d \mid \prod e \in C \mathbb{I}_{(\langle x \rangle - e)^T u \geq 0} = 1 \right\} \) is measurable. Then the sums in both the numerator and denominator are both measurable functions of \( u \), and the denominator cannot vanish. Therefore, \( \text{FracOptimal}(X \mid C, u) \) is a measurable function of \( u \).

Let \( g(X \mid u) := \text{FracOptimal}(X \mid C, u) \). Since \( \phi_i \cdot C = C \), lemma B.10 shows that \( g(X \mid u) \) is also invariant to joint permutation by \( \phi_i \). Since \( g \) is measurable and bounded \([0, 1]\), apply lemma B.5 to conclude that \( f(X \mid D_{\text{any}}) := \mathbb{E}_{u \sim D_{\text{any}}} [g(X \mid C, u)] \) is also invariant to joint permutation by \( \phi_i \).

Furthermore, if \( X' \subseteq X \subseteq C \), then \( f(X' \mid D_{\text{any}}) \leq f(X \mid D_{\text{any}}) \). So apply lemma B.9 to conclude that \( \text{FracOptimal}(B \mid C, D_{\text{any}}) := f(B \mid D_{\text{any}}) \geq f(A \mid D_{\text{any}}) := \text{FracOptimal}(A \mid C, D_{\text{any}}) \).

**Item 2.** Because \( X, C \) are finite sets, the denominator of \( \text{FracOptimal}(X \mid C, u) \) is never zero, and so the function is well-defined. \( \text{FracOptimal}(X \mid C, u) \) is an EU-determined function:

\[
\text{FracOptimal}(X \mid C, u) = \frac{\mathbb{I}_{x \in X \cap \{\arg\max_{e \in C} e^T u\}}}{\mathbb{I}_{\{\arg\max_{e \in C} e^T u\}}}
\]

Let \( \text{FracOptimal}(X \mid C, u) \) be the Boltzmann function:

\[
\text{Boltzmann}_T(X \mid C, u) = g_T \left( \left[ \begin{array}{c} V_X \\ x^T u \end{array} \right]_{x \in X}, \left[ \begin{array}{c} V_C \\ e^T u \end{array} \right]_{e \in C} \right)
\]

Therefore, by lemma B.11, \( \text{Boltzmann}_T(X \mid C, u) \) is invariant to joint permutation by \( \phi_i \).

Inspecting eq. (61), we see that \( g \) is continuous on \( u \) (and therefore measurable), and bounded \([0, 1]\) since \( X \subseteq C \) and the exponential function is positive. Therefore, by lemma B.5, the expectation version is also invariant to joint permutation for all permutations \( \phi \in S_d \): \( \text{Boltzmann}_T \left( \phi \cdot X \mid \phi \cdot C, \phi \cdot D_{\text{any}} \right) = \text{Boltzmann}_T \left( \phi \cdot X \mid \phi \cdot C, \phi \cdot D_{\text{any}} \right) \).

Since \( \phi_i \cdot C = C \), lemma B.10 shows that \( f(X \mid D_{\text{any}}) := \text{Boltzmann}_T(X \mid C, D_{\text{any}}) \) is also invariant under joint permutation by \( \phi_i \). Furthermore, if \( X' \subseteq X \), then \( f(X' \mid D_{\text{any}}) \leq \)}
We now show that best-of-\(k(X \mid C, u)\). Then apply lemma B.9 to conclude that Boltzmann\(_T(B \mid C, D_{\text{any}}) =: f(B \mid D_{\text{any}}) \overset{n_{\text{most}}}\geq f(A \mid D_{\text{any}}) =: \text{Boltzmann}\(_T(A \mid C, D_{\text{any}}).\)

**Item 5.** Let involution \(\phi \in S_d\) fix \(C\) (i.e. \(\phi \cdot C = C\)).

\[
\text{best-of-}k(X \mid C, u)
\]
\[
:= \mathbb{E}_{a_1, \ldots, a_k \sim \text{unif}(C)} \left[ \text{FracOptimal} \left( X \cap \{a_1, \ldots, a_k\} \mid \{a_1, \ldots, a_k\}, u \right) \right]
\]
\[
= \mathbb{E}_{\phi \sim \text{unif}(\phi C)} \left[ \text{FracOptimal} \left( (\phi \cdot X) \cap \{\phi \cdot a_1, \ldots, \phi \cdot a_k\} \mid \{\phi \cdot a_1, \ldots, \phi \cdot a_k\}, \phi \cdot u \right) \right]
\]
\[
= \mathbb{E}_{\phi \sim \text{unif}(\phi C)} \left[ \text{FracOptimal} \left( (\phi \cdot X) \cap \{\phi \cdot a_1, \ldots, \phi \cdot a_k\} \mid \{\phi \cdot a_1, \ldots, \phi \cdot a_k\}, \phi \cdot u \right) \right]
\]
\[
= \text{best-of-}k(\phi \cdot X \mid \phi \cdot C, \phi \cdot u).
\]

By the proof of item 2,

\[
\text{FracOptimal} \left( X \cap \{a_1, \ldots, a_k\} \mid \{a_1, \ldots, a_k\}, u \right) = \text{FracOptimal} \left( (\phi \cdot X) \cap \{\phi \cdot a_1, \ldots, \phi \cdot a_k\} \mid \{\phi \cdot a_1, \ldots, \phi \cdot a_k\}, \phi \cdot u \right)
\]

thus, eq. (64) holds. Since \(\phi \cdot C = C\) and since the distribution is uniform, eq. (65) holds. Therefore, best-of-\(k(X \mid C, u)\) is invariant to joint permutation by the \(\phi_i\), which are involutions fixing \(C\).

We now show that best-of-\(k(X \mid C, u)\) is measurable on \(u\).

\[
\text{best-of-}k(X \mid C, u)
\]
\[
:= \mathbb{E}_{a_1, \ldots, a_k \sim \text{unif}(C)} \left[ \text{FracOptimal} \left( X \cap \{a_1, \ldots, a_k\} \mid \{a_1, \ldots, a_k\}, u \right) \right]
\]
\[
= \frac{1}{|C|^k} \sum_{(a_1, \ldots, a_k) \in C^k} \text{FracOptimal} \left( X \cap \{a_1, \ldots, a_k\} \mid \{a_1, \ldots, a_k\}, u \right).
\]

Equation (69) holds because \(\text{FracOptimal}(X \mid C, u)\) is measurable on \(u\) by item 2, and measurable functions are closed under finite addition and scalar multiplication. Then best-of-\(k(X \mid C, u)\) is measurable on \(u\).

Let \(g(X \mid u) := \text{best-of-}k(X \mid C, u)\). Since \(\phi_i \cdot C = C\), lemma B.10 shows that \(g(X \mid u)\) is also invariant to joint permutation by \(\phi_i\). Since \(g\) is measurable and bounded \([0, 1]\), apply lemma B.5 to conclude that \(f(X \mid D_{\text{any}}) := \mathbb{E}_{u \sim D_{\text{any}}} g(X \mid C, u)\) is also invariant to joint permutation by \(\phi_i\).

Furthermore, if \(X' \subseteq X \subseteq C\), then \(f(X' \mid D_{\text{any}}) \leq f(X \mid D_{\text{any}})\). So apply lemma B.9 to conclude that best-of-\(k(B \mid C, D_{\text{any}}) =: f(B \mid D_{\text{any}}) \overset{n_{\text{most}}}\geq f(A \mid D_{\text{any}}) =: \text{best-of-}k(A \mid C, D_{\text{any}})\).

**Item 6.** Satisfice\(_t\) \((X \mid C, u)\) is an EU-determined function:

\[
\text{Satisfice}_t \left( X \mid C, u \right) = g_t \left( \begin{bmatrix} V_X & 0 \\ X \mathbf{u} \end{bmatrix} \right) \left( \begin{bmatrix} V_C \\ c \mathbf{u} \end{bmatrix} \right) \mathbf{1}_{e \in C}
\]
\[
= \frac{\sum_{v \in V_X} 1_{v \geq t} \mathbf{1}_{v \geq t}}{\sum_{v \in V_C} 1_{v \geq t} \mathbf{1}_{v \geq t}},
\]

with the function evaluating to 0 if the denominator is 0.

Then applying lemma B.11, Satisfice\(_t\) \((X \mid C, u)\) is invariant under joint permutation by the \(\phi_i\).

We now show that Satisfice\(_t\) \((X \mid C, u)\) is measurable on \(u\).

\[
\text{Satisfice}_t \left( X \mid C, u \right) = \begin{cases} \frac{\sum_{x \in X} 1_{x \in \{x' \in \mathcal{X}^t \mid x'^t \mathbf{u} \geq t\}}}{\sum_{e \in C} \sum_{x' \in \mathcal{X}^t \mid x'^t \mathbf{u} \geq t} 1_{e \in \mathcal{X}^t \mid x'^t \mathbf{u} \geq t}} & \exists e \in C : c^t \mathbf{u} \geq t, \\ 0 & \text{else.} \end{cases}
\]
Consider the two cases.

\[ \exists c \in C : c^T u \geq t \iff u \in \bigcup_{c \in C} \{ u' \in \mathbb{R}^d | c^T u \geq t \}. \]

The right-hand set is the union of finitely many halfspaces (which are measurable), and so the right-hand set is also measurable. Then the casing is a measurable function of \( u \). Clearly the zero function is measurable. Now we turn to the first case.

In the first case, eq. (72)’s indicator functions test each \( x, c \) for membership in a closed halfspace with respect to \( u \). Halfspaces are measurable sets. Therefore, the indicator function is a measurable function of \( u \), and so are the finite sums. Since the denominator does not vanish within the case, the first case as a whole is a measurable function of \( u \). Therefore, \( \text{Satis}_\ell(X | C, u) \) is measurable on \( u \).

Since \( \text{Satis}_\ell(X | C, u) \) is measurable and bounded \([0,1]\) (as \( X \subseteq C \)), apply lemma B.5 to conclude that \( \text{Satis}_\ell(X | C, D_{\text{any}}) = \text{Satis}_\ell(\phi \cdot X | \phi \cdot C, \phi \cdot D_{\text{any}}) \). Next, let \( f(X | D_{\text{any}}) := \text{Satis}_\ell(X | C, D_{\text{any}}) \). Since we just showed that \( \text{Satis}_\ell(X | C, D_{\text{any}}) \) is invariant to joint permutation by the involutions \( \phi_i \) and since \( \phi_i \cdot C = C, f(X | D_{\text{any}}) \) is also invariant to joint permutation by \( \phi_i \).

Furthermore, if \( X' \subseteq X \), we have \( f(X' | D_{\text{any}}) \leq f(X | D_{\text{any}}) \). Then applying lemma B.9, \( \text{Satis}_\ell(B | C, u) := f(B | D_{\text{any}}) \leq_{\text{most}} f(A | D_{\text{any}}) := \text{Satis}_\ell(A | C, u) \).

**Item 7.** Suppose \( P \) is uniform over \( C \) and consider any of the involutions \( \phi_i \).

\[
M_{q,p}(C | u) := \inf M \in \mathbb{R} \mid \mathbb{P}_{x \sim P} \left( x^T u > M \right) \leq q \quad (73)
\]

\[
= \inf \left\{ M \in \mathbb{R} \mid \mathbb{P}_{x \sim P} \left( (P_{\phi_\ell} x)^T (P_{\phi_\ell} u) > M \right) \leq q \right\} \quad (74)
\]

\[
= \inf \left\{ M \in \mathbb{R} \mid \mathbb{P}_{x \sim \phi_\ell^M} \left( x^T (P_{\phi_\ell} u) > M \right) \leq q \right\} \quad (75)
\]

\[
= \inf \left\{ M \in \mathbb{R} \mid \mathbb{P}_{x \sim P} \left( x^T (P_{\phi_\ell} u) > M \right) \leq q \right\} \quad (76)
\]

\[
=: M_{q,p}(\phi_i \cdot C | \phi_i \cdot u). \quad (77)
\]

Equation (74) follows by the orthogonality of permutation matrices. Equation (76) follows because if \( x \in \text{supp}(P) = C \), then \( \phi_i \cdot x \in C = \text{supp}(P) \), and furthermore \( P(x) = P(P_{\phi_\ell} x) \) by uniformity.

Now we show the invariance of \( C >_{M_{q,p}(C | u)} \) under joint permutation by \( \phi_i \):

\[
C >_{M_{q,p}(C|u)} := \left\{ c \in C \mid c^T u > M_{q,p}(C | u) \right\} \quad (78)
\]

\[
= \left\{ c \in C \mid (P_{\phi_\ell} c)^T (P_{\phi_\ell} u) > M_{q,p}(\phi_i \cdot C | \phi_i \cdot u) \right\} \quad (79)
\]

\[
= \left\{ c \in \phi_i \cdot C \mid c^T (P_{\phi_\ell} u) > M_{q,p}(\phi_i \cdot C | \phi_i \cdot u) \right\} \quad (80)
\]

\[
=: C >_{M_{q,p}(\phi_i \cdot C | \phi_i \cdot u)} \quad (81)
\]

Equation (79) follows by the orthogonality of permutation matrices and because \( M_{q,p}(C | u) = M_{q,p}(\phi_i \cdot C | \phi_i \cdot u) \) by eq. (77). A similar proof shows that \( C =_{M_{q,p}(C|u)} = C =_{M_{q,p}(\phi_i \cdot C | \phi_i \cdot u)} \).

Recall that

\[
Q_{q,p}(X | C, u) := \sum_{x \in X} \frac{P(x)}{q} \left( \mathbb{I}_{x \in C >_{M_{q,p}(C|u)}} + \frac{\mathbb{I}_{x \in C =_{M_{q,p}(C|u)}}}{P(C =_{M_{q,p}(C|u)})} \left( q - P\left( C >_{M_{q,p}(C|u)} \right) \right) \right). \quad (82)
\]

\[
Q_{q,p}(X | C, u) = Q_{q,p}(\phi_i \cdot X | \phi_i \cdot C, \phi_i \cdot u), \text{ since } Q \text{ is the sum of products of } \phi_i \text{-invariant quantities.}
\]
\( P(x) \) is non-negative because \( P \) is a probability distribution, and \( q \) is assumed positive. The indicator functions \( \mathbb{1} \) are non-negative. By the definition of \( M_q,p, P \left( C \triangleright M_q, p|C,u| \right) \leq q \). Therefore, eq. (82) is the sum of non-negative terms. Thus, if \( X' \subseteq X \), then \( Q_{q,p}(X' \mid C, u) \leq Q_{q,p}(X \mid C, u) \).

Let \( f(X \mid u) := Q_{q,p}(X \mid C, u) \). Since \( \phi_i \cdot C = C \) and since \( Q_{q,p}(X \mid C, u) = Q_{q,p}(\phi_i \cdot X \mid \phi_i \cdot C, \phi_i \cdot u) \), lemma B.10 shows that \( f(X \mid u) \) is also jointly invariant to permutation by \( \phi_i \). Lastly, if \( X' \subseteq X \), we have \( f(X' \mid D_{\text{any}}) \leq f(X \mid D_{\text{any}}) \).

Apply lemma B.9 to conclude that \( Q_{q,p}(B \mid C, u) := f(B \mid u) \geq \max_{\text{most} \in \mathbb{R}^5} f(A \mid u) := Q_{q,p}(A \mid C, u) \).

\[ \textbf{Conjecture B.13} \text{ (Orbit tendencies occur for more quantizer base distributions).} \] Proposition A.11’s item 7 holds for any base distribution \( P \) over \( C \) such that \( \min_{b \in B} P(b) \geq \max_{a \in A} P(a) \). Furthermore, \( Q_{q,p}(X \mid C, u) \) is measurable on \( u \) and so \( \max_{\text{most} \in \mathbb{R}^5} \) can be generalized to \( \geq \max_{\text{most} \in \mathbb{R}^5} \).

Appendix C Detailed analyses of MR scenarios

C.1 Action selection

Consider a bandit problem with five arms \( a_1, \ldots, a_5 \) partitioned \( A := \{a_1\}, B := \{a_2, \ldots, a_5\} \), which each action has a definite utility \( u_i \). There are \( T = 100 \) trials. Suppose the training procedure train uses the \( \epsilon \)-greedy strategy to learn value estimates for each arm. At the end of training, train outputs a greedy policy with respect to its value estimates. Consider any action-value initialization, and the learning rate is set \( \alpha := 1 \). To learn an optimal policy, at worst, the agent just has to try each action once.

\[ \textbf{Lemma C.1} \text{ (Lower bound on success probability of the train bandit).} \] Let \( u \in \mathbb{R}^5 \) assign strictly maximal utility to \( a_i \), and suppose train (described above) runs for \( T \geq 5 \) trials. Then \( p_{\text{train}}(\{a_i\} \mid u) \geq 1 - (1 - \frac{\epsilon}{2})^T \).

\[ \textit{Proof.} \] Since the trained policy can be stochastic,

\[ p_{\text{train}}(\{a_i\} \mid u) \geq \mathbb{P}(a_i \text{ is assigned probability 1 by the learned greedy policy}) . \]

Since \( a_i \) has strictly maximal utility which is deterministic, and since the learning rate \( \alpha := 1 \), if action \( a_i \) is ever drawn, it is assigned probability 1 by the learned policy. The probability that \( a_i \) is never explored is at most \( (1 - \frac{\epsilon}{2})^T \), because at worst, \( a_i \) is an ”explore” action (and not an ”exploit” action) at every time step, in which case it is ignored with probability \( 1 - \frac{\epsilon}{2} \).

\[ \textbf{Proposition C.2} \text{ (The train bandit is 4-retradatable).} \] \( p_{\text{train}} \) is \( (\mathbb{R}^5, A \xrightarrow{4} B) \)-tradatable.

\[ \textit{Proof.} \] Let \( \phi_i := a_1 \leftrightarrow a_i \) for \( i = 2, \ldots, 5 \) and let \( \Theta := \mathbb{R}^5 \). We want to show that whenever \( u \in \mathbb{R}^5 \) induces \( p_{\text{train}}(A \mid u) > p_{\text{train}}(B \mid u) \), retargeting \( u \) will get train to instead learn to pull a \( B \)-action:

\[ p_{\text{train}}(A \mid \phi_i \cdot u) < p_{\text{train}}(B \mid \phi_i \cdot u) . \]

Suppose we have such a \( u \). If \( u \) is constant, a symmetry argument shows that each action has equal probability of being selected, in which case \( p_{\text{train}}(A \mid u) = \frac{1}{5} < \frac{1}{4} = p_{\text{train}}(B \mid u) \)—a contradiction. Therefore, \( u \) is not constant. Similar symmetry arguments show that \( A \)-action \( a_1 \) has strictly maximal utility (\( u_1 > \max_{i=2,\ldots,5} u_i \)).

But for \( T = 100 \), lemma C.1 shows that \( p_{\text{train}}(A \mid u) = p_{\text{train}}(\{a_1\} \mid u) \approx 1 \) and \( p_{\text{train}}(\{a_{i \neq 1}\} \mid u) \approx 0 \), the converse statement holds when considering \( \phi_i \cdot u \) instead of \( u \). Therefore, train satisfies definition 3.5’s item 1 (tradatability). These \( \phi_i \cdot u \in \Theta := \mathbb{R}^5 \) because \( \mathbb{R}^5 \) is closed under permutation by \( S_5 \), satisfying item 2.

Consider another \( u' \in \mathbb{R}^5 \) such that \( p_{\text{train}}(A \mid u') > p_{\text{train}}(B \mid u') \), and consider \( i \neq j \). By the above symmetry arguments, \( u' \) must also assign \( a_1 \) maximal utility. By lemma C.1, \( p_{\text{train}}(\{a_i\} \mid \phi_i \cdot u) \approx 1 \) and \( p_{\text{train}}(\{a_j\} \mid \phi_j \cdot u) \approx 0 \) since \( i \neq j \), and vice versa when considering \( \phi_j \cdot u \) instead of \( \phi_i \cdot u \). Then since \( \phi_i \cdot u \) and \( \phi_j \cdot u \) induce distinct probability distributions over learned actions, they cannot be the same utility function. This satisfies item 3.
Corollary C.3 (The train bandit has orbit-level tendencies). \( p_{train}(B \mid u) \gtrsim^{A} \mathbb{R}^{+} p_{train}(A \mid u) \).


C.2 Observation reward maximization

Let \( T \) be a reasonably long rollout length, so that \( O_{T\text{-reach}} \) is large—many different step-\( T \) observations can be induced.

Proposition C.4 (Final reward maximization has strong orbit-level incentives in MR). Let \( n := \left\lfloor \frac{|O_{leave}|}{|O_{stay}|} \right\rfloor \). \( \max_{\text{most: } \mathbb{R}^{+}} p_{\max}(O_{leave} \mid R) \gtrsim^{A} \max_{\text{most: } \mathbb{R}^{+}} p_{\max}(O_{stay} \mid R) \).

Proof. Consider the vector space representation of observations, \( \mathbb{R}^{|O|} \). Define \( A := \{ e_{o} \mid o \in O_{stay} \} \), \( B := \{ e_{o} \mid o \in O_{leave} \} \), and \( C := O_{T\text{-reach}} = A \cup B \) the union of \( O_{stay}, O_{leave} \).

Since \( |O_{leave}| \gtrsim^{A} |O_{stay}| \) by assumption that \( T \) is reasonably large, consider the involution \( \phi_{1} \in S_{|O|} \) which embeds \( O_{stay} \) into \( O_{leave} \), while fixing all other observations. If possible, produce another involution \( \phi_{2} \) which also embeds \( O_{stay} \) into \( O_{leave} \), which fixes all other observations, and which “doesn’t interfere with \( \phi_{1} \)” (i.e. \( \phi_{2} \cdot (\phi_{1} \cdot A) = \phi_{1} \cdot A \)). We can produce \( n := \left\lfloor \frac{|O_{leave}|}{|O_{stay}|} \right\rfloor \) such involutions.

Therefore, \( B \) contains \( n \) copies (definition A.7) of \( A \) via involutions \( \phi_{1}, \ldots, \phi_{n} \). Furthermore, \( \phi_{i} \cdot (A \cup B) = A \cup B \), since each \( \phi_{i} \) swaps \( A \) with \( B' \subseteq B \), and fixes all \( b \in B \setminus B' \) by assumption. Thus, \( \phi \cdot C = C \).

By proposition A.11’s item 2, \( \max_{\text{most: } \mathbb{R}^{+}} \frac{\text{FracOptimal}(B \mid C, R)}{\text{FracOptimal}(A \mid C, R)} \gtrsim^{A} \max_{\text{most: } \mathbb{R}^{+}} \frac{\text{FracOptimal}(X \mid C, R)}{\text{FracOptimal}(X \mid C, R)} \) Since \( p_{\max} \) uniformly randomly chooses a maximal-reward observation to induce, \( \forall X \subseteq C : p_{\max}(X \mid R) = \frac{\text{FracOptimal}(X \mid C, R)}{\text{FracOptimal}(X \mid C, R)} \). Therefore, \( \max_{\text{most: } \mathbb{R}^{+}} p_{\max}(O_{leave} \mid R) \gtrsim^{A} \max_{\text{most: } \mathbb{R}^{+}} p_{\max}(O_{stay} \mid R) \).

We want to reason about the probability that \( \text{decide} \) leaves the initial room by time \( T \) in its rollout trajectories.

\[
p_{\text{decide}}(\text{leave} \mid \theta) := \mathbb{P}_{\pi \sim \text{decide}(\theta), \tau \sim \pi_{s_{0}}} (\tau \text{ has left the first room by step } T), \quad (83)
\]
\[
p_{\text{decide}}(\text{stay} \mid \theta) := \mathbb{P}_{\pi \sim \text{decide}(\theta), \tau \sim \pi_{s_{0}}} (\tau \text{ has not left the first room by step } T). \quad (84)
\]
We want to show that reward maximizers tend to leave the room: \( p_{\text{max}}(\text{leave} \mid R) \geq \max_{\theta \in \Theta} p_{\text{max}}(\text{stay} \mid R) \). However, we must be careful: In general, \( p_{\text{max}}(O_{\text{leave}} \mid R) \neq p_{\text{max}}(\text{leave} \mid R) \) and \( p_{\text{max}}(O_{\text{stay}} \mid R) \neq p_{\text{max}}(\text{stay} \mid R) \). For example, suppose that \( o_T \in O_{\text{leave}} \). By the definition of \( O_{\text{leave}} \), \( o_T \) can only be observed if the agent has left the room by time step \( T \), and so the trajectory \( \tau \) must have left the first room. The converse argument does not hold: The agent could leave the first room, re-enter, and then wait until time \( T \). Although one of the doors would have been opened (fig. 2), the agent can also open the door without leaving the room, and then realize the same step-\( T \) observation. Therefore, this observation doesn’t belong to \( O_{\text{leave}} \).

**Lemma C.5** (Room-status inequalities for MR).

\[
\begin{align*}
p_{\text{decide}}(\text{stay} \mid \theta) &\leq p_{\text{decide}}(O_{\text{stay}} \mid \theta), \quad \text{(85)} \\
\text{and } p_{\text{decide}}(O_{\text{leave}} \mid \theta) &\leq p_{\text{decide}}(\text{leave} \mid \theta). \quad \text{(86)}
\end{align*}
\]

**Proof.** For any decide,

\[
\begin{align*}
p_{\text{decide}}(\text{stay} \mid \theta) &= \mathbb{P} \left( \tau \text{ stays through step } T \right) \quad \text{(87)} \\
&= \sum_{o \in O} \mathbb{P} \left( o \text{ at step } T \mid \tau \right) \mathbb{P} \left( \tau \text{ stays } \mid o \text{ at step } T \right) \quad \text{(88)} \\
&= \sum_{o \in O_{\text{reach}}} \mathbb{P} \left( o \text{ at step } T \right) \mathbb{P} \left( \tau \text{ stays } \mid o \text{ at step } T \right) \quad \text{(89)} \\
&= \sum_{o \in O_{\text{stay}}} \mathbb{P} \left( o \text{ at step } T \right) \mathbb{P} \left( \tau \text{ stays } \mid o \text{ at step } T \right) \quad \text{(90)} \\
&\leq \sum_{o \in O_{\text{stay}}} \mathbb{P} \left( o \text{ at step } T \right) \quad \text{(91)} \\
&= \mathbb{P} \left( o_T \in O_{\text{stay}} \right) \quad \text{(92)} \\
&= p_{\text{decide}}(O_{\text{stay}} \mid \theta). \quad \text{(93)}
\end{align*}
\]

Equation (90) holds because the definition of \( O_{\text{reach}} \) ensures that if \( o \notin O_{\text{reach}} \), \( \tau \mid o \theta \) = 0. Because \( o \in O_{\text{reach}} \setminus O_{\text{stay}} \) implies that \( \tau \) left and so

\[
\mathbb{P} \left( \tau \text{ stays } \mid o \text{ at step } T \right) = 0,
\]

eq. (91) follows. Then we have shown eq. (85).

For eq. (86),

\[
\begin{align*}
p_{\text{decide}}(O_{\text{leave}} \mid \theta) &= \mathbb{P} \left( \tau \in O_{\text{leave}} \right) \quad \text{(95)} \\
&= \sum_{o \in O_{\text{leave}}} \mathbb{P} \left( o \text{ at step } T \right) \quad \text{(96)} \\
&= \sum_{o \in O_{\text{leave}}} \mathbb{P} \left( o \text{ at step } T \right) \mathbb{P} \left( \tau \text{ leaves } \mid o \text{ at step } T \right) \quad \text{(97)} \\
&= \sum_{o \in O_{\text{leave}}} \mathbb{P} \left( o \text{ at step } T \right) \mathbb{P} \left( \tau \text{ leaves } \mid o \text{ at step } T \right) \quad \text{(98)} \\
&= \sum_{o \in O} \mathbb{P} \left( o \text{ at step } T \right) \mathbb{P} \left( \tau \text{ leaves } \mid o \text{ at step } T \right) \quad \text{(99)} \\
&= \mathbb{P} \left( \tau \text{ has left the first room by step } T \right) \quad \text{(100)}
\end{align*}
\]
We will now conduct a more detailed analysis and conclude that only a sword (by leaving the room, some realizable trajectory induces \( R \subseteq (\text{FracOptimal} \ Y) \) and then uniformly randomly chooses a policy which implements procedure takes \( \alpha \). Suppose all feature coefficients \( \in \) are plausible. Then we have shown eq. (86).

\( \square \)

**Corollary C.6** (Final reward maximizers tend to leave the first room in MR).

\[
p_{\text{max}}(\text{leave} \mid R) \geq_{\text{most: } R \circ} p_{\text{max}}(\text{stay} \mid R).
\]

**Proof.** Using lemma C.5 and proposition C.4, apply lemma B.1 with \( f_0(R) := p_{\text{max}}(\text{leave} \mid R) \); \( f_1(R) := p_{\text{max}}(\text{O}_\text{leave} \mid R) \), \( f_2(R) := p_{\text{max}}(\text{O}_\text{stay} \mid R) \), \( f_3(R) := p_{\text{max}}(\text{stay} \mid R) \) to conclude that \( p_{\text{max}}(\text{leave} \mid R) \geq_{\text{most: } R \circ} p_{\text{max}}(\text{stay} \mid R) \).

\( \square \)

### C.3 Featurized reward maximization

\( \Theta := \mathbb{R}^O \) assumes we will specify complicated reward functions over observations, with \(|O|\) degrees of freedom in their specification. Any observation can get any number. However, reward functions are often specified more compactly. For example, in section 4.3, the (additively) featurized reward function \( R_{\text{feat}}(o_T) := \text{feat}(o_T)^\top \alpha \) has four degrees of freedom. Compared to typical reward functions (which would look like “random noise” to a human), \( R_{\text{feat}} \) more easily trains competent policies because of the regularities between the reward and the state features.

In this setup, \( p_{\text{max}} \) chooses a policy which induces a step-\( T \) observation with maximal reward. Reward depends only on the feature vector of the final observation—more specifically, on the agent’s item counts. There are more possible item counts available by first leaving the room, than by staying.

We will now conduct a more detailed analysis and conclude that \( p_{\text{max}}(\text{O}_\text{leave} \mid \alpha) \geq_{\text{most: } R \circ} p_{\text{max}}(\text{O}_\text{stay} \mid \alpha) \). Informally, we can retarget which items the agent prioritizes, and thereby retarget from \( \text{O}_\text{stay} \to \text{O}_\text{leave} \).

Consider the featurization function which takes as input an observation \( o \in O \):

\[
\text{feat}(o) := \begin{pmatrix}
\text{# of keys in inventory shown by } o \\
\text{# of swords in inventory shown by } o \\
\text{# of torches in inventory shown by } o \\
\text{# of amulets in inventory shown by } o
\end{pmatrix}.
\]

Consider \( A_{\text{feat}} := \{ \text{feat}(o) \mid o \in \text{O}_\text{stay} \} \), \( B_{\text{feat}} := \{ \text{feat}(o) \mid o \in \text{O}_\text{leave} \} \).

Let \( e_i \in \mathbb{R}^4 \) be the standard basis vector with a 1 in entry \( i \) and 0 elsewhere. When restricted to the room shown in fig. 2, the agent can either acquire the key in the first room and retain it until step \( T \) (\( e_1 \)), or reach time step \( T \) empty-handed (\( 0 \)). We conclude that \( A_{\text{feat}} = \{ e_1, 0 \} \).

For \( B_{\text{feat}} \), recall that in section 4.2 we assumed the rollout length \( T \) to be reasonably large. Then by leaving the room, some realizable trajectory induces \( o_T \) displaying an inventory containing only a sword (\( e_2 \)), or only a torch (\( e_3 \)), or only an amulet (\( e_4 \)), or nothing at all (\( 0 \)). Therefore, \( \{ e_2, e_3, e_4, 0 \} \subseteq B_{\text{feat}} \). \( B_{\text{feat}} \) contains 3 copies of \( A_{\text{feat}} \) (definition A.7) via involutions \( \phi_i : 1 \leftrightarrow i, i \neq 1 \). Suppose all feature coefficient vectors \( \alpha \in \mathbb{R}^4 \) are plausible. Then \( \Theta := \mathbb{R}^4 \).

Let us be more specific about what is entailed by featurized reward maximization. The \( \text{decide}_{\text{max}}(\alpha) \) procedure takes \( \alpha \) as input and then considers the reward function \( o \mapsto \text{feat}(o)^\top \alpha \). Then, \( \text{decide}_{\text{max}} \) uniformly randomly chooses an observation \( o_T \in O_T \text{-reach} \) which maximizes this featurized reward, and then uniformly randomly chooses a policy which implements \( o_T \).

**Lemma C.7** (FracOptimal inequalities). Let \( X \subseteq Y' \subseteq Y \subseteq \mathbb{R}^d \) be finite, and let \( u \in \mathbb{R}^d \). Then

\[
\text{FracOptimal} \left( X \mid Y, u \right) \leq \text{FracOptimal} \left( X \mid Y', u \right) \leq \text{FracOptimal} \left( X \cup (Y \setminus Y') \mid Y, u \right).
\]

**Proof.** For finite \( X_1 \subseteq \mathbb{R}^d \), let \( \text{Best}(X_1 \mid u) := \arg \max_{x_1 \in X_1} x_1^\top u \). Suppose \( y' \in \text{Best}(Y' \mid u) \), but \( y' \notin \text{Best}(Y \mid u) \). Then for all \( \alpha \in \text{Best}(Y' \mid u) \),

\[
\alpha^\top u = y'^\top u < \max_{y \in Y} y^\top u.
\]
Then eq. (106) follows since \(\mathcal{Y}_0 \not\subseteq \mathcal{Y}'\). Then in this case, eq. (106) follows because \(\mathcal{Y}_0 \subseteq \mathcal{Y}'\). On the other hand, if \(\mathcal{Y}_0 \not\subseteq \mathcal{Y}'\), then since \(\mathcal{Y}_0 \subseteq \mathcal{Y}'\), we have \(\mathcal{Y}_0 \cap \mathcal{Y}' = \emptyset\). Then eq. (106) follows, and eq. (107) is non-negative. Either way, eq. (107)’s inequality holds. To show the second inequality, we handle the two cases separately.

**Subset case.** Suppose that \(\mathcal{Y}_0 \subseteq \mathcal{Y}'\).

\[
\frac{\mathcal{Y}_0 \cap \mathcal{Y}_0}{\mathcal{Y}_0} \leq \frac{\mathcal{Y}_0 \cap \mathcal{Y}_0}{\mathcal{Y}_0} + \frac{\mathcal{Y}_0 \setminus \mathcal{Y}_0}{\mathcal{Y}_0} + \frac{\mathcal{Y}_0 \setminus \mathcal{Y}_0}{\mathcal{Y}_0}
\]

(108)

Equation (108) follows because when \(n \leq d, k \geq 0\), we have \(\frac{n}{d} \leq \frac{n+k}{d+k}\). For eq. (110), since \(\mathcal{Y}_0 \subseteq \mathcal{Y}'\), we must have

\[
\mathcal{Y}_0 = \mathcal{Y}_0 \setminus \mathcal{Y}_0 \cup \mathcal{Y}_0 \setminus \mathcal{Y}_0.
\]

But then

\[
\mathcal{Y}_0 \cap \mathcal{Y}_0 = \left(\mathcal{Y}_0 \setminus \mathcal{Y}_0 \cap \mathcal{Y}_0 \setminus \mathcal{Y}_0\right) \cup \left(\mathcal{Y}_0 \setminus \mathcal{Y}_0 \cap \mathcal{Y}_0 \setminus \mathcal{Y}_0\right)
\]

(115)

Then eq. (110) follows. Equation (111) follows since

\[
\mathcal{Y}_0 = \mathcal{Y}_0 \setminus \mathcal{Y}_0 \cup \mathcal{Y}_0 \setminus \mathcal{Y}_0.
\]

Equation (112) follows since \(\mathcal{Y}_0 \subseteq \mathcal{Y}'\), and so

\[
\mathcal{Y}_0 \cap \mathcal{Y}_0 = \mathcal{Y}_0 \cap \mathcal{Y}_0.
\]

Equation (113) follows because \(\mathcal{Y}_0 \subseteq \mathcal{Y}'\) is disjoint of \(\mathcal{Y}_0 \setminus \mathcal{Y}_0\). We have shown that

\[
\mathcal{Y}_0 \subseteq \mathcal{Y}_0 \setminus \mathcal{Y}_0 \cap \mathcal{Y}_0 \setminus \mathcal{Y}_0
\]

in this case.
**Disjoint case.** Suppose that \( \text{Best} (Y' | u) \cap \text{Best} (Y | u) = \emptyset \).

\[
\left\lvert \text{Best} (Y' | u) \cap X \right\rvert \leq 1 \tag{117}
\]

\[
= \left\lvert \text{Best} (Y \setminus Y' | u) \right\rvert \tag{118}
\]

\[
= \frac{\text{Best} (Y \setminus Y' | u) \cap (Y \setminus Y')}{\text{Best} (Y \setminus Y' | u)} \tag{119}
\]

\[
= \frac{\text{Best} (Y \setminus Y' | u) \cap (X \cup (Y \setminus Y'))}{\text{Best} (Y \setminus Y' | u)} \tag{120}
\]

\[
= \frac{\text{Best} (Y | u) \cap (X \cup (Y \setminus Y'))}{\text{Best} (Y | u)} \tag{121}
\]

\[
= \text{FracOptimal} (X \cup (Y \setminus Y') | Y, u) \tag{122}
\]

Equation (117) follows because \( \text{Best} (Y' | u) \cap X \subseteq \text{Best} (Y' | u) \). For eq. (120), note that we trivially have \( \text{Best} (Y' | u) \cap \text{Best} (Y \setminus Y' | u) = \emptyset \), and also that \( X \subseteq Y' \). Therefore, \( \text{Best} (Y \setminus Y' | u) \cap X = \emptyset \), and eq. (121) follows. Finally, the disjointness assumption implies that

\[
\max_{y^T u \in Y''} y^T u < \max_{y^T u \in Y} y^T u.
\]

Therefore, the optimal elements of \( Y \) must come exclusively from \( Y \setminus Y' \); i.e. \( \text{Best} (Y | u) = \text{Best} (Y \setminus Y' | u) \). Then eq. (121) follows, and we have shown that

\[
\text{FracOptimal} (X | Y', u) \leq \text{FracOptimal} (X \cup (Y \setminus Y') | Y, u)
\]

in this case. \(\square\)

**Conjecture C.8** (Generalizing lemma C.7). Lemma C.7 and Turner et al. [2021]'s Lemma E.26 have extremely similar functional forms. How can they be unified?

**Proposition C.9** (Featurized reward maximizers tend to leave the first room in MR).

\[
p_{\max} (\text{leave} | \alpha) \geq_{\text{most}} \mathbb{R}^4 p_{\max} (\text{stay} | \alpha). \tag{123}
\]

**Proof.** We want to show that \( p_{\max} (O_{\text{leave}} | \alpha) \geq_{\text{most}} \mathbb{R}^4 p_{\max} (O_{\text{stay}} | \alpha) \). Recall that \( A_{\text{feat}} = \{e_1, 0\} \), \( B'_{\text{feat}} := \{e_2, e_3, e_4\} \subseteq B_{\text{feat}} \).

\[
p_{\max} (\text{stay} | \alpha) \tag{124}
\]

\[
\leq p_{\max} (O_{\text{stay}} | \alpha) \tag{125}
\]

\[
:= \mathbb{P}_{\pi \sim \text{decide}_{\text{max}} (\alpha), \tau \sim \pi | s_0} (o_T \in O_{\text{stay}}) \tag{126}
\]

\[
\leq \text{FracOptimal} (\{e_1\} | C_{\text{feat}}, \alpha) + \mathbb{P}_{\pi \sim \text{decide}_{\text{max}} (\alpha), \tau \sim \pi | s_0} (o_T \in O_{\text{stay}}, \text{feat}(o_T) = 0) \tag{127}
\]

\[
\leq \text{FracOptimal} (\{e_1\} | \{e_1, e_2, e_3, e_4\}, \alpha) + \mathbb{P}_{\pi \sim \text{decide}_{\text{max}} (\alpha), \tau \sim \pi | s_0} (o_T \in O_{\text{stay}}, \text{feat}(o_T) = 0) \tag{128}
\]

31
we conclude that 

\[ \text{FracOptimal}\left(\left\{ \{e_2, e_3, e_4\} \mid \{e_1, e_2, e_3, e_4\} , \alpha \right\} \right) \]

\[ + P \pi \sim \text{decide}_{\max(\alpha)}, \quad (o_T \in O_{\text{leave}} \text{, } \text{feat}(o_T) = 0) \]

\[ \leq \text{FracOptimal}\left(\left\{ \{e_2, e_3, e_4\} \cup \left(\text{C} \text{feat} \setminus \{e_1, e_2, e_3, e_4\}\right) \right\} \mid \text{C} \text{feat}, \alpha \right) \]

\[ = \text{FracOptimal}\left(\text{C} \text{feat} \setminus \{e_1\} \mid \text{C} \text{feat}, \alpha \right) \]

\[ \leq P \pi \sim \text{decide}_{\max(\alpha)}, \quad (o_T \in O_{\text{leave}}) \]

\[ =: p_{\max}(\text{O}_{\text{leave}} \mid \alpha) \]

\[ \leq p_{\max}(\text{leave} \mid \alpha). \]

Equation (124) and eq. (135) hold by lemma C.5. If \( o_T \in O_{\text{stay}} \) is realized by \( p_{\max} \) and \( \text{feat}(o_T) \neq 0 \), then we must have \( \text{feat}(o_T) = \{e_1\} \) be optimal and so the \( e_1 \) inventory configuration is realized. Therefore, eq. (128) follows. Equation (129) follows by applying the first inequality of lemma C.7 with \( X := \{e_1\} \), \( Y' := \{e_1, e_2, e_3, e_4\} \), \( Y := \text{C} \text{feat} \).

By applying proposition A.11’s item 2 with \( A := A_{\text{feat}} = \{e_1\} \), \( B' := B'_{\text{feat}} = \{e_2, e_3, e_4\} \), \( C' := A \cup B' \), we have

\[ \text{FracOptimal}\left(\left\{ \{e_1\} \mid \{e_1, e_2, e_3, e_4\} , \alpha \right\} \right) \leq_{\text{most}} P \pi \sim \text{decide}_{\max(\alpha)}, \quad \text{FracOptimal}\left(\left\{ \{e_2, e_3, e_4\} \mid \{e_1, e_2, e_3, e_4\} , \alpha \right\} \right). \]

Furthermore, observe that

\[ P \pi \sim \text{decide}_{\max(\alpha)}, \quad (o_T \in O_{\text{stay}} \text{, } \text{feat}(o_T) = 0) \leq P \pi \sim \text{decide}_{\max(\alpha)}, \quad (o_T \in O_{\text{leave}} \text{, } \text{feat}(o_T) = 0) \]

because either 0 is not optimal (in which case both sides equal 0), or else 0 is optimal, in which case the right side is strictly greater. This can be seen by considering how \( \text{decide}_{\max(\alpha)} \) uniformly randomly chooses an observation in which the agent ends up with an empty inventory. As argued previously, the vast majority of such observations can only be induced by leaving the first room.

Combining eq. (136) and eq. (137), eq. (130) follows. Equation (131) follows by applying the second inequality of lemma C.7 with \( X := \{e_2, e_3, e_4\} \), \( Y' := \{e_1, e_2, e_3, e_4\} \), \( Y := \text{C} \text{feat} \). If \( \text{feat}(o_T) \in B_{\text{feat}} \) is realized by \( p_{\max} \), then by the definition of \( B_{\text{feat}} \), \( o_T \in O_{\text{leave}} \) is realized, and so eq. (133) follows.

Then by applying lemma B.1 with

\[ f_0(\alpha) := p_{\max}(\text{leave} \mid \alpha), \]

\[ f_1(\alpha) := \text{FracOptimal}\left(\left\{ \{e_1\} \mid \{e_1, e_2, e_3, e_4\} , \alpha \right\} \right), \]

\[ f_2(\alpha) := \text{FracOptimal}\left(\left\{ \{e_2, e_3, e_4\} \mid \{e_1, e_2, e_3, e_4\} , \alpha \right\} \right), \]

\[ f_3(\alpha) := p_{\max}(\text{stay} \mid \alpha), \]

we conclude that

\[ p_{\max}(\text{leave} \mid \alpha) \geq_{\text{most}} P \pi \sim \text{decide}_{\max(\alpha)}, \quad p_{\max}(\text{stay} \mid \alpha). \]

Lastly, note that if \( 0 \in \Theta \) and \( f(A \mid 0) > f(B \mid 0) \), \( f \) cannot be even be simply retargetable for the \( \Theta \) parameter set. This is because \( \forall \phi \in S_4, \phi \cdot 0 = 0 \). For example, inductive bias ensures that, absent a reward signal, learned policies tend to stay in the initial room in MR. This is one reason why section 4.3’s analysis of the policy tendencies of reinforcement learning excludes the all-zero reward function.

C.4 Reasoning for why DQN can’t explore well

In section 4.3, we wrote:

Mnih et al. [2015]’s DQN isn’t good enough to train policies which leave the first room of MR, and so DQN (trivially) cannot be retargetable away from the first room.
via the reward function. There isn’t a single featurized reward function for which DQN visits other rooms, and so we can’t have \( \alpha \) such that \( \phi \cdot \alpha \) retargets the agent to \( O_{\text{leave}} \). DQN isn’t good enough at exploring.

We infer this is true from Nair et al. [2015], which shows that vanilla DQN gets zero score in MR. Thus, DQN never even gets the first key. Thus, DQN only experiences state-action-state transitions which didn’t involve acquiring an item, since (as shown in fig. 3) the other items are outside of the first room, which requires a key to exit. In our analysis, we considered a reward function which is featurized over item acquisition.

Therefore, for all pre-key-acquisition state-action-state transitions, the featurized reward function returns exactly the same reward signals as those returned in training during the published experiments (namely, zero, because DQN can never even get to the key in order to receive a reward signal). That is, since DQN only experiences state-action-state transitions which didn’t involve acquiring an item, and the featurized reward functions only reward acquiring an item, it doesn’t matter what reward values are provided upon item acquisition—DQN’s trained behavior will be the same. Thus, a DQN agent trained on any featurized reward function will not explore outside of the first room.

### Appendix D Lower bounds on MDP power-seeking incentives for optimal policies

Turner et al. [2021] prove conditions under which at least half of the orbit of every reward function incentivizes power-seeking behavior. For example, in fig. 4, they prove that avoiding \( \emptyset \) maximizes average per-timestep reward for at least half of reward functions. Roughly, there are more self-loop states (\( \ell, \ell \downarrow, \ell \uparrow, r \downarrow, r \uparrow \)) available if the agent goes left or right instead of up towards \( \emptyset \). We strengthen this claim, with corollary D.12 showing that for at least three-quarters of the orbit of every reward function, it is average-optimal to avoid \( \emptyset \).

Therefore, we answer Turner et al. [2021]’s open question of whether increased number of environmental symmetries quantitatively strengthens the degree to which power-seeking is incentivized. The answer is yes. In particular, it may be the case that only one in a million state-based reward functions makes it average-optimal for Pac-Man to die immediately.

![Figure 4: A toy MDP for reasoning about power-seeking tendencies. Reproduced from Turner et al. [2021].](image)

We will briefly restate several definitions needed for our key results, theorem D.11 and corollary D.12. For explanation, see Turner et al. [2021].

**Definition D.1** (Non-dominated linear functionals). Let \( X \subseteq \mathbb{R}^{|S|} \) be finite. \( \text{ND}(X) := \left\{ \mathbf{x} \in X \mid \exists \mathbf{r} \in \mathbb{R}^{|S|} : \mathbf{x}^\top \mathbf{r} > \max_{\mathbf{x}^\prime \in X \setminus \{\mathbf{x}\}} \mathbf{x}^\prime \top \mathbf{r} \right\} \).

**Definition D.2** (Bounded reward function distribution). \( \mathcal{D}_{\text{bound}} \) is the set of bounded-support probability distributions \( \mathcal{D}_{\text{bound}} \).

**Remark.** When \( n = 1 \), lemma D.3 reduces to the first part of Turner et al. [2021]’s lemma E.24, and lemma D.5 reduces to the first part of Turner et al. [2021]’s lemma E.28.

**Lemma D.3** (Quantitative expectation superiority lemma). Let \( A, B \subseteq \mathbb{R}^d \) be finite and let \( g : \mathbb{R} \to \mathbb{R} \) be a (total) increasing function. Suppose \( B \) contains \( n \) copies of \( \text{ND}(A) \). Then

\[
E_{\mathbf{r} \sim \mathcal{D}_{\text{bound}}} \left[ g \left( \max_{\mathbf{b} \in B} \mathbf{b}^\top \mathbf{r} \right) \right] \geq_{\text{most}} n D_{\text{bound}} E_{\mathbf{r} \sim \mathcal{D}_{\text{bound}}} \left[ g \left( \max_{\mathbf{a} \in A} \mathbf{a}^\top \mathbf{r} \right) \right].
\]  

(142)
Proof. Because $g : \mathbb{R} \rightarrow \mathbb{R}$ is increasing, it is measurable (as is max).

Let $L := \inf_{r \in \text{supp}(D_{\text{bound}})} \max_{x \in X} x^\top r$, $U := \sup_{r \in \text{supp}(D_{\text{bound}})} \max_{x \in X} x^\top r$. Both exist because $D_{\text{bound}}$ has bounded support. Furthermore, since $g$ is monotone increasing, it is bounded $[g(L), g(U)]$ on $[L, U]$. Therefore, $g$ is measurable and bounded on each $\text{supp}(D_{\text{bound}})$, and so the relevant expectations exist for all $D_{\text{bound}}$.

For finite $X \subseteq \mathbb{R}^d$, let $f(X \mid u) := g(\max_{x \in X} x^\top u)$. By lemma B.11, $f$ is invariant under joint permutation by $S_d$. Furthermore, $f$ is measurable because $g$ and max are. Therefore, apply lemma B.5 to conclude that $f(X \mid D_{\text{bound}}) := \mathbb{E}_{u \sim D_{\text{bound}}} [g(\max_{x \in X} x^\top u)]$ is also invariant under joint permutation by $S_d$ (with $f$ being bounded when restricted to $\text{supp}(D_{\text{bound}})$). Lastly, if $X' \subseteq X$, $f(X' \mid D_{\text{bound}}) \leq f(X \mid D_{\text{bound}})$ because $g$ is increasing.

\[
\mathbb{E}_{u \sim D_{\text{bound}}} \left[ g \left( \max_{a \in A} a^\top u \right) \right] = \mathbb{E}_{u \sim D_{\text{bound}}} \left[ g \left( \max_{a \in \text{ND}(A)} a^\top u \right) \right] \leq \text{most: } D_{\text{any}} \mathbb{E}_{\tau \sim D_{\text{bound}}} \left[ g \left( \max_{b \in B} b^\top \tau \right) \right].
\]

Equation (143) follows by corollary E.11 of Turner et al., 2021. Equation (144) follows by applying lemma B.9 with $f$ as defined above with the $\phi_1, \ldots, \phi_n$ guaranteed by the copy assumption.

Definition D.4 (Linear functional optimality probability [Turner et al., 2021]). For finite $A, B \subseteq \mathbb{R}^{|S|}$, the probability under $D_{\text{any}}$ that $A$ is optimal over $B$ is

\[
p_{D_{\text{any}}} (A \geq B) := \mathbb{P}_{\tau \sim D_{\text{any}}} \left( \max_{a \in A} a^\top \tau \geq \max_{b \in B} b^\top \tau \right).
\]

Lemma D.5 (Quantitative optimality probability superiority lemma). Let $A, B, C \subseteq \mathbb{R}^d$ be finite and let $Z$ satisfy ND $(C) \subseteq Z \subseteq C$. Suppose that $B$ contains $n$ copies of ND $(A)$ via involutions $\phi_i$. Furthermore, let $B_{\text{extra}} := B \setminus \bigcup_{i=1}^n \phi_i \cdot \text{ND}(A)$; suppose that for all $i$, $\phi_i \cdot (Z \setminus B_{\text{extra}}) = Z \setminus B_{\text{extra}}$.

Then $p_{D_{\text{any}}} (B \geq C) \geq n \text{most: } D_{\text{any}} p_{D_{\text{any}}} (A \geq C)$.

Proof. For finite $X, Y \subseteq \mathbb{R}^d$, let

\[
g(X, Y \mid D_{\text{any}}) := p_{D_{\text{any}}} (X \geq Y) = \mathbb{E}_{D_{\text{any}}} \left[ \mathbb{1}_{\max_{x \in X} x^\top u \geq \max_{y \in Y} y^\top u} \right].
\]

By the proof of item 1 of proposition A.11, $g$ is the expectation of a $u$-measurable function. $g$ is an EU function, and so lemma B.11 shows that it is invariant to joint permutation by $\phi_1$. Letting $f_Y(X \mid D_{\text{any}}) := g(X, Y \mid D_{\text{any}})$, lemma B.10 shows that $f_Y(X \mid D_{\text{any}}) = f_Y(\phi_1 \cdot X \mid \phi_1 \cdot D_{\text{any}})$ whenever the $\phi_1$ satisfy $\phi_1 \cdot Y = Y$.

Furthermore, if $X' \subseteq X$, then $f_Y(X' \mid D_{\text{any}}) \leq f_Y(X \mid D_{\text{any}})$.

\[
p_{D_{\text{any}}} (A \geq C) = p_{D_{\text{any}}} (\text{ND}(A) \geq C) \leq p_{D_{\text{any}}} (\text{ND}(A) \geq Z \setminus B_{\text{extra}}) \leq p_{D_{\text{any}}} (B \geq Z \setminus B_{\text{extra}}) \leq p_{D_{\text{any}}} (B \cup B_{\text{extra}} \geq Z) = p_{D_{\text{any}}} (B \geq Z) = p_{D_{\text{any}}} (B \geq C).
\]

Equation (145) follows by Turner et al., 2021’s lemma E.12’s item 2 with $X := A$, $X' := \text{ND}(A)$ (similar reasoning holds for $C$ and $Z$ in eq. (150)). Equation (146) follows by the first inequality of lemma E.26 of [Turner et al., 2021] with $X := A, Y := C, Y' := Z \setminus B_{\text{extra}}$. Equation (147) follows by applying lemma B.9 with the $f_{Z \setminus B_{\text{extra}}}$ defined above. Equation (148) follows by the second inequality of lemma E.26 of [Turner et al., 2021] with $X := A, Y := Z, Y' := Z \setminus B_{\text{extra}}$. Equation (149) follows because $B_{\text{extra}} \subseteq B$.  

34
Letting $f_0(D_{\text{any}}) := p_{D_{\text{any}}} (A \geq C)$, $f_1(D_{\text{any}}) := p_{D_{\text{any}}} (ND (A) \geq Z \setminus B_{\text{extra}})$, $f_2(D_{\text{any}}) := p_{D_{\text{any}}} (B \geq Z \setminus B_{\text{extra}})$, $f_3(D_{\text{any}}) := p_{D_{\text{any}}} (B \geq C)$, apply lemma B.1 to conclude that

$$p_{D_{\text{any}}} (A \geq C) \leq_p \max_{D_{\text{any}}} p_{D_{\text{any}}} (B \geq C).$$

\[\square\]

**Definition D.6** (Rewardless MDP [Turner et al., 2021]). $(S, A, T)$ is a rewardless MDP with finite state and action spaces $S$ and $A$, and stochastic transition function $T : S \times A \to \Delta(S)$. We treat the discount rate $\gamma$ as a variable with domain $[0, 1]$. Let $D : S \times A \to \Delta(S)$ be the standard basis vector

$$P_{\text{any}} = \text{average-optimal policy set for reward function } R \text{ is } \Pi^R (R) := \left\{ \pi \in \Pi \mid \forall s \in S : d^{\pi, s} \in \arg \max_{d \in \text{RSD} (s)} d^{\pi, s} \right\} \text{ (the policies which induce optimal RSDs at all states). For } D \subseteq \text{RSD} (s), \text{ the average optimality probability is } P_{D_{\text{any}}} (D, \text{average}) := P_{R \sim D_{\text{any}}} \left( \exists d^{\pi, s} \in D : \pi \in \Pi^R (R) \right).$$

**Remark.** Theorem D.11 generalizes the first claim of Turner et al. [2021]’s theorem 6.13, and corollary D.12 generalizes the first claim of Turner et al. [2021]’s corollary 6.14.

**Theorem D.11** (Quantitatively, average-optimal policies tend to end up in “larger” sets of RSDs). Let $D', D \subseteq \text{RSD} (s)$. Suppose that $D$ contains $n$ copies of $D'$ and that the sets $D' \cup D$ and $\text{RSD} (s) \setminus (D' \cup D)$ have pairwise orthogonal vector elements (i.e. pairwise disjoint vector support). Then $P_{D_{\text{any}}} (D', \text{average}) \geq_\text{most: } D_{\text{any}} P_{D_{\text{any}}} (D, \text{average}).$

**Proof.** Let $D_i := \phi_i \cdot D'$, where $D_i \subseteq D$ by assumption. Let $S := \left\{ s' \in S \mid \max_{d \in D'} d^{\phi_i, s'} > 0 \right\}$. Define $\phi_i (s') := \begin{cases} \phi_i (s') & \text{if } s' \in S \\ s' & \text{else.} \end{cases}$ (151)

Since $\phi_i$ is an involution, $\phi_i'$ is also an involution. Furthermore, $\phi_i' \cdot D' = D_i$, $\phi_i' \cdot D_j = D_i$, and $\phi_i' \cdot D_j = D_j$ for $j \neq i$ because we assumed that these equalities hold for $\phi_i$, and $D', D_i, D_j \subseteq D' \cup D$ and so the vectors of these sets have support contained in $S$.

Let $D'^* := D' \cup_{i=1}^n D_i \cup \left( \text{RSD} (s) \setminus (D' \cup D) \right)$. By an argument mirroring that in the proof of theorem 6.13 in Turner et al. [2021] and using the fact that $\phi_i' \cdot D_j = D_j$ for all $i \neq j$, $\phi_i' \cdot D'^* = D'^*$. Consider $Z := (\text{RSD} (s) \setminus (D' \cup D)) \cup D' \cup D$. First, $Z \subseteq \text{RSD} (s)$ by definition. Second, $\text{RSD} (s) = \text{RSD} (s) \setminus (D' \cup D) \cup (\text{RSD} (s) \cap D') \cup (\text{RSD} (s) \cap D) \subseteq Z$. Note that $D'^* = Z \setminus (D \setminus \bigcup_{i=1}^n D_i)$. $P_{D_{\text{any}}} (D', \text{average}) = P_{D_{\text{any}}} (D' \geq \text{RSD} (s))$ (152)

$$\leq_\text{most: } D_{\text{any}} P_{D_{\text{any}}} (D \geq \text{RSD} (s))$$ (153)

$$= P_{D_{\text{any}}} (D, \text{average}).$$ (154)

Since $\phi_i' \cdot D' \subseteq D$ and $\text{ND} (D') \subseteq D'$, $\phi_i' \cdot \text{ND} (D') \subseteq D$ and so $D$ contains $n$ copies of $\text{ND} (D')$ via involutions $\phi_i'$. Then eq. (153) holds by applying lemma D.5 with $A := D'$, $B_i := D_i$ for all $i = 1, \ldots, n$, $B := D, C := \text{RSD} (s), Z$ as defined above, and involutions $\phi_i'$ which satisfy $\phi_i' \cdot (Z \setminus (B \setminus \bigcup_{i=1}^n B_i)) = \phi_i' \cdot D'^* = D'^* = Z \setminus (B \setminus \bigcup_{i=1}^n B_i)$. \[\square\]
Corollary D.12 (Quantitatively, average-optimal policies tend not to end up in any given 1-cycle).
Let $D' := \{e_{s_1'}, \ldots, e_{s_k'}\}$, $D_r := \{e_{s_1}, \ldots, e_{s_n,k}\} \subseteq \text{RSD}(s)$ be disjoint, for $n \geq 1, k \geq 1$. Then
$$\mathbb{P}_{D_{\text{any}}} (D', \text{average}) \leq_{\text{most: } D_{\text{any}}} \mathbb{P}_{D_{\text{any}}} (\text{RSD}(s) \setminus D', \text{average}).$$

Proof. For each $i \in \{1, \ldots, n\}$, let
$$\phi_i := (s'_1 \ s_{(i-1)\cdot k+1}) \cdots (s'_{k} \ s_{(i-1)\cdot k+k}),$$
$$D_i := \{e_{s_{(i-1)\cdot k+1}}, \ldots, e_{s_{(i-1)\cdot k+k}}\},$$
$$D := \text{RSD}(s) \setminus D'.$$
Each $D_i \subseteq D_r \subseteq \text{RSD}(s) \setminus D'$ by disjointness of $D'$ and $D_r$.

$D$ contains $n$ copies of $D'$ via involutions $\phi_1, \ldots, \phi_n$. $D' \cup D = \text{RSD}(s)$ and $\text{RSD}_{nd}(s) \setminus \text{RSD}(s) = \emptyset$ trivially have pairwise orthogonal vector elements.

Apply theorem D.11 to conclude that
$$\mathbb{P}_{D_{\text{any}}} (D', \text{average}) \leq_{n} \mathbb{P}_{D_{\text{any}}} (\text{RSD}(s) \setminus D', \text{average}).$$

Let $A := \{e_1, e_2\}, B \subseteq \mathbb{R}^5, C := A \cup B$. Conjecture D.13 conjectures that e.g.
$$\mathbb{P}_{D'} (B \geq C) \geq_{3} \mathbb{P}_{D'} (A \geq C).$$

Conjecture D.13 (Fractional quantitative optimality probability superioriy lemma). Let $A, B, C \subseteq \mathbb{R}^d$ be finite. If $A = \bigcup_{j=1}^{n} A_j$ and $\bigcup_{i=1}^{m} B_i \subseteq B$ such that for each $A_j, B$ contains $n$ copies $(B_1, \ldots, B_n)$ of $A_j$ via involutions $\phi_{ji}$ which also fix $\phi_{ji} \cdot A_j = A_j$ for $j' \neq j$, then
$$\mathbb{P}_{D_{\text{any}}} (B \geq C) \geq_{n} \mathbb{P}_{D_{\text{any}}} (A \geq C).$$

We suspect that any proof of the conjecture should generalize lemma B.7 to the fractional set copy containment case.