Universal Agent Mixtures and the Geometry of Intelligence

Samuel Allen Alexander
The U.S. Securities and Exchange Commission

David Quarel
Australian National University

Len Du
WooliesX

Marcus Hutter
DeepMind & Australian National University

Abstract

Inspired by recent progress in multi-agent Reinforcement Learning (RL), in this work we examine the collective intelligent behaviour of theoretical universal agents by introducing a weighted mixture operation. Given a weighted set of agents, their weighted mixture is a new agent whose expected total reward in any environment is the corresponding weighted average of the original agents' expected total rewards in that environment. Thus, if RL agent intelligence is quantified in terms of performance across environments, the weighted mixture's intelligence is the weighted average of the original agents' intelligences. This operation enables various interesting new theorems that shed light on the geometry of RL agent intelligence, namely: results about symmetries, convex agent-sets, and local extrema. We also show that any RL agent intelligence measure based on average performance across environments, subject to certain weak technical conditions, is identical (up to a constant factor) to performance within a single environment dependent on said intelligence measure.

1 INTRODUCTION

Multi-agent Reinforcement Learning (or multi-agent RL) (Weiss, 1993; Littman, 1994; Zhang et al., 2021; Hernández-Orallo et al., 2011), as with other flavors of RL, has been enjoying increased attention in artificial intelligence research (Lanctot et al., 2017). The most obvious way to conceive multi-agent RL, is to passively consider the collective behavior of aggregated intelligent agents. In fact, multi-agent RL was first introduced as "collective learning" (Weiss, 1993) even before being explicitly identified as multi-agent reinforcement learning (Littman, 1994).

One significant recent trend in machine learning (beyond just RL) is Federated Reinforcement Learning (Chen et al., 2021), which is mainly about programs physically running on disparate devices collaborating to form a more powerful artificial intelligence. This approach conceptually borrows from how humans collaborate. In this regard, RL had a much earlier head start with Feudal Reinforcement Learning (Dayan and Hinton, 1992), which borrows concepts which have been around for many hundreds of years in human social organization, and on which research remains active today (Johnson and Dana, 2020). Multi-agent methods have also received considerable attention for their usage in highly complex real-time video-games (Vinyals et al., 2019) (building off of success in simpler games like Atari games (Mnih et al., 2015) and Go (Silver et al., 2017)). RL is no stranger to collaboration (Kok and Vlassis, 2006) or to cooperation (Qiu et al., 2021), including even collaboration with human users (Li et al., 2021).

In this work, we draw inspiration from *sortition*, which is yet another way of social organization. In sortition, instead of the whole citizenry collaborating on individual decisions, citizens are chosen *by lottery* and granted temporary power. Thus in a statistical sense, each citizen enjoys a certain amount of total expected power. The roots of sortition trace back to the original Athenian democracy (Hansen, 1991), and sortition has attracted recent scientific curiosity, and even advocacy in real-world governance (Flanigan et al., 2021; Sintomer, 2018; Bouricius, 2013).

More specifically, we examine the expected intelligence of a single combined mixture agent formed from a group of agents using sortition. Given agents $\vec{\pi} = (\pi_1, \dots, \pi_n)$, imagine an agent σ who, at the start of each agent-environment interaction, randomly chooses an agent π_i to act as for that entire interaction (that is, the selection only occurs *once*, at the beginning, and then persists—we do *not*

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

mean that one of the π_i is randomly chosen on every single turn). Imagine that each candidate π_i is so chosen with probability (or *weight*) w_i (where $w_1 + \cdots + w_n = 1$). If each π_i would get total expected reward R_i from an environment, we would expect σ to get total expected reward $\vec{w} \cdot \vec{R} = w_1 R_1 + \cdots + w_n R_n$ from that environment.

Starting from practical (e.g., conformant with OpenAI Gym (Brockman et al., 2016)) implementations of agents $\pi_1, \ldots, \pi_n$ as above, σ could easily be implemented as a new agent who, upon instantiation, uses a random number generator to determine which π_i to act as, and stores that decision in internal memory. But this sort of construction is not possible in more abstract, theoretical RL frameworks such as (Legg and Hutter, 2007, 2005), where agents are mathematical functions which take histories as input and output action-space probability-distributions. Such functions have no "instantiation", no access to a true random number generator, and no concept of "internal memory". A key result of ours is that nevertheless, it is possible to define an agent σ having the same exact performance as the above-described sortition agent, entirely within an abstract, theoretical RL framework. This is important because the constraints of the theoretical framework facilitate rigorous mathematical proofs of properties of σ. To see that the construction is non-trivial, consider instead an agent ρ who, upon instantiation, examines the computer's system clock, and determines to play as π_1 if the clock says "AM" or as π_2 if the clock says "PM". Such a ρ could certainly be implemented in an OpenAI gym conformant way, but clearly has no counterpart in a formal RL framework with no notion of a system clock.

The fact that the weighted mixture agent σ's expected total reward in any environment is the corresponding weighted average of the expected total rewards of $\pi_1, \ldots, \pi_n$ will allow us to prove multiple interesting results that shed light on the geometry of RL agent performance and performance-based intelligence measures. We obtain the following results and applications:

- (Section 3) By guaranteeing that "the expected reward of a weighted mixture is the weighted average of the expected rewards", we establish a method of combining agents without the risk of unforeseen side-effects. For example, if several agents have different weaknesses, then, a priori, one might worry that, combining those agents, those weaknesses might compound each other, leading to a combined weakness larger than the sum of the individual weaknesses. Our mixture agent construction avoids this, as well as other emergent behavior which would violate the above quote.
- (Section 4) We consider two different ways an intelligence measure can be symmetric with respect to the operation of interchanging rewards and punishments. We prove that these two symmetry notions are equivalent. This has implications in the search for inherently desirable properties of universal Turing machines.
- (Section 5) We introduce notions of *discernability* and *separability* of sets of RL agents, and characterize the latter in terms of the former and closure under our mixture operation. If agents are thought of as points in space, then these properties are analogous to higher-dimensional convexity notions from convex geometry. These results can help determine what sort of things can or cannot be incentivized in RL, in a formal sense (similar to using the Pumping Lemma to show that certain languages are not regular).
- (Section 6) We introduce a notion of an agent being a strict local extremum of an intelligence measure, and we show that any such agent is, in a certain formal sense, deterministic. This result is highly applicable in the quest to optimize RL agents, as it implies that nothing is gained by allowing agents to invoke genuine random number generators, e.g. expensive RNGs based on quantum mechanics, etc.
- (Section 7) Finally, we use our technique to mix environments, rather than agents. Using the resulting mixture environments, we prove that every intelligence measure satisfying certain properties is necessarily equivalent (up to a constant multiple) to performance in some particular environment.

2 PRELIMINARIES

Throughout the paper, we implicitly fix non-empty finite sets $\mathcal{A}$ of *actions*, $\mathcal{O}$ of *observations*, and $\mathcal{R} \subseteq \mathbb{Q} \cap [-1, 1]$ of *rewards*. By ε we mean the empty sequence. By $\mathcal{E}$ we mean $\mathcal{O} \times \mathcal{R}$ (the set of all observation-reward pairs); elements of $\mathcal{E}$ are called *percepts*. By $\Delta\mathcal{A}$ (resp. $\Delta\mathcal{E}$) we mean the set of all probability distributions on $\mathcal{A}$ (resp. on $\mathcal{E}$).

Definition 1. *(Agents, environments, etc.)*

1. *We denote the set of all finite sequences of alternating percept-action pairs $x_1 y_1 \ldots x_t y_t$ by $(\mathcal{E}\mathcal{A})^*$. We also include ε in $(\mathcal{E}\mathcal{A})^*$. Nonempty elements of $(\mathcal{E}\mathcal{A})^*$ have the form $x_1 y_1 \ldots x_t y_t$ where each x_i is a percept and each y_i is an action.*
2. *We denote the set of all sequences of the form sx (where $s \in (\mathcal{E}\mathcal{A})^*$, $x \in \mathcal{E}$, and sx is the result of appending x to s) by $(\mathcal{E}\mathcal{A})^*\mathcal{E}$. Elements of $(\mathcal{E}\mathcal{A})^*\mathcal{E}$ of length > 1 have the form $x_1 y_1 \ldots x_{t-1} y_{t-1} x_t$ (each x_i a percept, each y_i an action).*
3. *An* agent[1] *is a function $\pi : (\mathcal{E}\mathcal{A})^*\mathcal{E} \to \Delta\mathcal{A}$. For any $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$, we write $\pi(\cdot|h)$ for the value of π at h,*

[1]Not to be confused with a *policy*, which would simply be a function $\mathcal{O} \to \Delta\mathcal{A}$.

and for any $y \in \mathcal{A}$, we write $\pi(y|h)$ for $(\pi(\cdot|h))(y)$. Intuitively, for any action y, $\pi(y|h)$ is the probability that agent π takes action y in response to history h.

4. *An* environment *is a function $\mu : (\mathcal{EA})^* \to \Delta\mathcal{E}$. For every $h \in (\mathcal{EA})^*$, we write $\mu(\cdot|h)$ for the value of μ at h, and for any $x \in \mathcal{E}$, we write $\mu(x|h)$ for $(\mu(\cdot|h))(x)$. If $x = (o, r)$ ($o \in \mathcal{O}$, $r \in \mathcal{R}$), we may also write $\mu(o, r|h)$ for $(\mu(\cdot|h))(x)$. Intuitively, $\mu(o, r|h)$ is the probability that environment μ issues percept (o, r) (observation o and reward r) to the agent in response to history h.*

Remark 2. *Note that in Definition 1 part 3, we require, e.g., $\pi(\cdot|x_1y_1x_2)$ to be defined even if $\pi(y_1|x_1) = 0$, in which case the initial percept-action sequence $x_1y_1x_2$ would have probability 0 of ever occurring in any agent-environment interaction. Intuitively: an agent must choose actions even in response to histories that would never occur with nonzero probability. This convention, in which we follow Legg and Hutter (2007), simplifies many definitions.*

Definition 3. *By $\mathcal{H}$ we mean $((\mathcal{EA})^*) \cup ((\mathcal{EA})^*\mathcal{E})$, in other words, $\mathcal{H}$ is the set of alternating percept-action sequences that are empty or else start with a percept and can end with either a percept or an action. We refer to elements h of $\mathcal{H}$ as* histories *(a history may terminate with either a percept or an action).*

Definition 4. *For all agents π, histories h, and environments μ, we define real numbers $P^\pi(h)$, $P_\mu(h)$, and $P^\pi_\mu(h)$ inductively as follows.*

- *If $h = \varepsilon$ then $P^\pi(h) = P_\mu(h) = P^\pi_\mu(h) = 1$.*
- *If $h = gx$ (some $x \in \mathcal{E}$) then $P^\pi(h) = P^\pi(g)$, $P_\mu(h) = P_\mu(g)\mu(x|g)$, and $P^\pi_\mu(h) = P^\pi_\mu(g)\mu(x|g)$.*
- *If $h = gy$ (some $y \in \mathcal{A}$) then $P^\pi(h) = P^\pi(g)\pi(y|g)$, $P_\mu(h) = P_\mu(g)$, and $P^\pi_\mu(h) = P^\pi_\mu(g)\pi(y|g)$.*

Intuitively: $P^\pi(h)$ is the conditional probability π will choose the actions in h assuming the environment which π is interacting with chooses the percepts in h; $P_\mu(h)$ is the conditional probability μ will choose the percepts in h assuming the agent which μ is interacting with chooses the actions in h; and $P^\pi_\mu(h)$ is the probability that π and μ will choose h's actions and percepts when interacting together.

Some authors, such as Hutter (2009), would write $P(h)$ or a variation thereof for P^π_μ, if π and μ are clear from context.

One could alternately more directly define

$$\begin{aligned} &P^\pi(x_1y_1 \ldots x_ty_t) \\ &= \pi(y_1|x_1)\pi(y_2|x_1y_1x_2)\cdots\pi(y_t|x_1y_1\ldots x_t), \end{aligned}$$

and similarly define $P^\pi(x_1y_1 \ldots x_t)$, and likewise for P_μ and for P^π_μ.

Lemma 5. *For all h, π, μ as in Definition 4,*

$$P^\pi_\mu(h) = P^\pi(h)P_\mu(h).$$

Proof. See Supplementary Materials. □

In the following definition (and the rest of the paper), $\mathbb{N}$ denotes the set $\{0, 1, 2, \ldots\}$ of non-negative integers.

Definition 6. *(Performance in an environment) Let π be an agent, μ an environment.*

1. *For every $t \in \mathbb{N}$, we define*

$$V^\pi_{\mu,t} = \sum_{h \in X_t} R(h)P^\pi_\mu(h)$$

where $X_t \subseteq \mathcal{H}$ is the set of all length-$2t$ histories (i.e., all $h \in \mathcal{H}$ of the form $x_1y_1 \ldots x_ty_t$ (each $x_i \in \mathcal{E}$, each $y_i \in \mathcal{A}$) provided $t > 0$) and $R(h)$ is the sum of the rewards in h. Intuitively, $V^\pi_{\mu,t}$ is the expected total reward if π were to interact with μ for t steps. Note that $X_0 = \{\varepsilon\}$ and so $V^\pi_{\mu,0} = 0$.

2. *We define $V^\pi_\mu = \lim_{t\to\infty} V^\pi_{\mu,t}$, provided the limit converges to a real number. Intuitively, V^π_μ is the expected total reward which π would extract from μ.*

Note that it is possible for V^π_μ to be undefined. For example, if μ is an environment which always issues reward $(-1)^t$ in response to the agent's tth action y_t, then V^π_μ is undefined for every agent π. We will only be interested in environments μ such that V^π_μ is always defined. Note also that, following Legg and Hutter (2007), we delegate any possible reward discounting to the environments themselves, rather than build a fixed reward discounting factor into the definition of V^π_μ.

Definition 7. *An environment μ is* well-behaved *if the following requirements hold: $\mu(x|h) \in \mathbb{Q}$ for all $x \in \mathcal{E}$ and $h \in (\mathcal{EA})^*$; μ is Turing computable; and for every agent π, V^π_μ exists and $-1 \leq V^\pi_\mu \leq 1$. Let W be the set of all well-behaved environments.*

Definition 8. *By a* weighted intelligence measure*, we mean a function $\Upsilon : (\Delta\mathcal{A})^{(\mathcal{EA})^*\mathcal{E}} \to \mathbb{R}$ (where $(\Delta\mathcal{A})^{(\mathcal{EA})^*\mathcal{E}}$ denotes the set of all agents) such that there exist non-negative reals $\{w_\mu\}_{\mu\in W}$ such that the following condition holds: for every agent π, $\Upsilon(\pi) = \sum_{\mu\in W} w_\mu V^\pi_\mu$.*

The prototypical weighted intelligence measure is the Legg-Hutter intelligence measure Υ introduced by Legg and Hutter (2007), where each well-behaved μ is weighed using the universal prior (Li and Vitányi, 2008; Hutter, 2003), i.e., given weight $2^{-K(\mu)}$ where K denotes Kolmogorov complexity ($K(\mu)$ exists because of the Turing computability requirement in Definition 7). This depends on a background universal Turing machine, the choice of which is highly nontrivial (Leike and Hutter, 2015).

3 MIXTURE AGENTS

Before defining mixture agents, we will first extend some of the above definitions to vectors of agents.

Definition 9. *Suppose $\vec{\pi} = (\pi_1, \ldots, \pi_n)$ is a vector of agents, μ is an environment, $h \in \mathcal{H}$, $t \in \mathbb{N}$, and Υ is a weighted intelligence measure. We define:*

- $P^{\vec{\pi}}(h) = (P^{\pi_1}(h), \ldots, P^{\pi_n}(h))$.
- $P^{\vec{\pi}}_\mu(h) = (P^{\pi_1}_\mu(h), \ldots, P^{\pi_n}_\mu(h))$.
- $V^{\vec{\pi}}_{\mu,t} = (V^{\pi_1}_{\mu,t}, \ldots, V^{\pi_n}_{\mu,t})$.
- $V^{\vec{\pi}}_\mu = (V^{\pi_1}_\mu, \ldots, V^{\pi_n}_\mu)$, *if $V^{\pi_1}_\mu, \ldots, V^{\pi_n}_\mu$ are defined.*
- $\Upsilon(\vec{\pi}) = (\Upsilon(\pi_1), \ldots, \Upsilon(\pi_n))$.

Definition 10. *If $\vec{u} = (u_1, \ldots, u_n)$ and $\vec{v} = (v_1, \ldots, v_n)$ are any two equal-length vectors of real numbers, then their* dot product *is defined to be $\vec{u} \cdot \vec{v} = u_1 v_1 + \cdots + u_n v_n$.*

Now we are ready to define mixture agents. As a motivating example, suppose we want to combine two agents into a joint agent, but we are worried that doing so might lead to unexpected emergent behavior. For example, we do not want the two agents' weaknesses to compound each other and give rise to a joint weakness larger than the sum of the two agents' weaknesses. We will show below that the following construction avoids such unexpected behavior.

Definition 11. *(Mixture agents) Suppose $\vec{\pi} = (\pi_1, \ldots, \pi_n)$ are agents and $\vec{w} = (w_1, \ldots, w_n)$ are positive real numbers with $w_1 + \cdots + w_n = 1$. Define the* mixture agent *$\vec{w} \cdot \vec{\pi}$ as follows: for all $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$, $y \in \mathcal{A}$, let*

$$(\vec{w} \cdot \vec{\pi})(y|h) = \begin{cases} \dfrac{\vec{w} \cdot P^{\vec{\pi}}(hy)}{\vec{w} \cdot P^{\vec{\pi}}(h)} & \text{if } \vec{w} \cdot P^{\vec{\pi}}(h) \neq 0, \\ 1/|\mathcal{A}| & \text{otherwise.} \end{cases}$$

Remark 12. *Definition 11 is complicated by the way Definition 1 part 3 forces us to include the case $(\vec{w} \cdot \vec{\pi})(y|h) = 1/|\mathcal{A}|$ when $\vec{w} \cdot P^{\vec{\pi}}(h) = 0$ (see Remark 2): we are obligated to specify how $\vec{w} \cdot \vec{\pi}$ chooses actions even in response to "impossible" histories for $\vec{w} \cdot \vec{\pi}$ (histories containing actions $\vec{w} \cdot \vec{\pi}$ would never take in those circumstances).*

Intuitively, $\vec{w} \cdot \vec{\pi}$ can be thought of as being driven by an entity who believes that actions are to be chosen by one of the $\vec{\pi}$, but does not know which. The entitiy initially assigns each w_i to π_i as a prior, and attempts to guess the probability of each action being chosen by the unknown agent π_i, using these priors to do so. As new actions are seen, said priors are updated using Bayes' rule.

Lemma 13. *If $\vec{\pi}$ and $\vec{w}$ are as in Definition 11 then the mixture agent $\vec{w} \cdot \vec{\pi}$ is an agent (per Definition 1 part 3).*

Proof. See Supplementary Materials. $\square$

We will frequently use Lemma 13 without explicit mention. For example, the lemma allows us to speak of $P^{\vec{w}\cdot\vec{\pi}}(h)$ (Definition 4), $V^{\vec{w}\cdot\vec{\pi}}_\mu$ (Definition 6), etc., and we will freely do so without explicitly citing Lemma 13.

Theorem 14. *(Commutativity of $\vec{w}$) Let $\vec{\pi} = (\pi_1, \ldots, \pi_n)$ be agents. Let $\vec{w} = (w_1, \ldots, w_n)$ be positive reals with $w_1 + \cdots + w_n = 1$. Let μ be any environment. Then:*

1. *For any $h \in \mathcal{H}$, $P^{\vec{w}\cdot\vec{\pi}}(h) = \vec{w} \cdot P^{\vec{\pi}}(h)$.*
2. *For any $h \in \mathcal{H}$, $P^{\vec{w}\cdot\vec{\pi}}_\mu(h) = \vec{w} \cdot P^{\vec{\pi}}_\mu(h)$.*
3. *For any $t \in \mathbb{N}$, $V^{\vec{w}\cdot\vec{\pi}}_{\mu,t} = \vec{w} \cdot V^{\vec{\pi}}_{\mu,t}$.*
4. *("The expected reward of a weighted mixture is the weighted average of the expected rewards") If $V^{\vec{\pi}}_\mu$ is defined, then $V^{\vec{w}\cdot\vec{\pi}}_\mu = \vec{w} \cdot V^{\vec{\pi}}_\mu$.*
5. *("The intelligence of a weighted mixture is the weighted average of the intelligences") For any weighted intelligence measure Υ, $\Upsilon(\vec{w}\cdot\vec{\pi}) = \vec{w}\cdot\Upsilon(\vec{\pi})$.*

Proof. (1) By induction on h.

Case 1: $h = \varepsilon$. Then

$$P^{\vec{w}\cdot\vec{\pi}}(h) = 1 = w_1 \cdot 1 + \cdots + w_n \cdot 1 = \vec{w} \cdot P^{\vec{\pi}}(h).$$

Case 2: $h = gx$ for some $x \in \mathcal{E}$. Then

$$\begin{aligned} P^{\vec{w}\cdot\vec{\pi}}(h) &= P^{\vec{w}\cdot\vec{\pi}}(g) && \text{(Definition 4)} \\ &= \vec{w} \cdot P^{\vec{\pi}}(g) && \text{(Induction)} \\ &= \vec{w} \cdot P^{\vec{\pi}}(gx) = \vec{w} \cdot P^{\vec{\pi}}(h). && \text{(Definition 4)} \end{aligned}$$

Case 3: $h = gy$ for some $y \in \mathcal{A}$.

Subcase 3.1: $P^{\vec{w}\cdot\vec{\pi}}(g) = 0$. By induction $\vec{w} \cdot P^{\vec{\pi}}(g) = 0$. Since the w_i are positive, this implies each $P^{\pi_i}(g) = 0$. Thus each

$$\begin{aligned} w_i P^{\pi_i}(gy) &= w_i P^{\pi_i}(g) \pi_i(y|g) && \text{(Definition 4)} \\ &= 0 w_i \pi_i(y|g) = 0, \end{aligned}$$

i.e., $\vec{w} \cdot P^{\vec{\pi}}(gy) = 0$. And

$$\begin{aligned} P^{\vec{w}\cdot\vec{\pi}}(gy) &= P^{\vec{w}\cdot\vec{\pi}}(g)(\vec{w} \cdot \vec{\pi})(y|g) && \text{(Definition 4)} \\ &= 0(\vec{w} \cdot \vec{\pi})(y|g) = 0, \end{aligned}$$

so $P^{\vec{w}\cdot\vec{\pi}}(h) = \vec{w} \cdot P^{\vec{\pi}}(h) = 0$.

Subcase 3.2: $P^{\vec{w}\cdot\vec{\pi}}(g) \neq 0$. Then

$$\begin{aligned} P^{\vec{w}\cdot\vec{\pi}}(h) &= P^{\vec{w}\cdot\vec{\pi}}(g)(\vec{w} \cdot \vec{\pi})(y|g) && \text{(Definition 4)} \\ &= P^{\vec{w}\cdot\vec{\pi}}(g) \frac{\vec{w} \cdot P^{\vec{\pi}}(gy)}{\vec{w} \cdot P^{\vec{\pi}}(g)} && \text{(Definition 11)} \\ &= \vec{w} \cdot P^{\vec{\pi}}(g) \frac{\vec{w} \cdot P^{\vec{\pi}}(gy)}{\vec{w} \cdot P^{\vec{\pi}}(g)} && \text{(Induction)} \\ &= \vec{w} \cdot P^{\vec{\pi}}(gy) = \vec{w} \cdot P^{\vec{\pi}}(h). && \text{(Basic Algebra)} \end{aligned}$$

(2) Follows from (1) and Lemma 5.

(3) With X_t and R as in Definition 6, we compute:

$$
\begin{aligned}
V_{\mu,t}^{\vec{w}\cdot\vec{\pi}} &= \textstyle\sum_{h\in X_t} R(h)P_\mu^{\vec{w}\cdot\vec{\pi}}(h) && \text{(Def. 6)}\\
&= \textstyle\sum_{h\in X_t} \vec{w}\cdot R(h)P_\mu^{\vec{\pi}}(h) && \text{(By (2))}\\
&= \vec{w}\cdot \textstyle\sum_{h\in X_t} R(h)P_\mu^{\vec{\pi}}(h) && \text{(Vect. algebra)}\\
&= \vec{w}\cdot \left(\textstyle\sum_{h\in X_t} R(h)P_\mu^{\pi_i}(h)\right)_{i=1}^n && \text{(Def. 9 part 2)}\\
&= \vec{w}\cdot V_{\mu,t}^{\vec{\pi}}. && \text{(Def. 9 part 3)}
\end{aligned}
$$

(4) Follows from (3) and Definition 6 part 2.

(5) Follows from (4) and Definition 8. ☐

Parts 4–5 of Theorem 14 show our mixture operation (Definition 11) avoids unexpected emergent behavior. Mixing two agents in this way cannot result in a joint agent whose weaknesses (or strengths) exceed the summed weaknesses (or strengths) of the original two agents, at least as far as measured by expected performance in RL environments.

4 EQUIVALENCE OF WEAK AND STRONG SYMMETRY

In this section, we will investigate two symmetry properties which a weighted intelligence measure might satisfy. A priori, one property seems stricly stronger, but we will show that in fact, they are equivalent. Throughout this section, we assume that the background set $\mathcal{R}$ has the following additional property: whenever $\mathcal{R}$ contains any reward r, then $\mathcal{R}$ also contains $-r$.

Definition 15. *(Dual Agents)*

1. *For each $h \in \mathcal{H}$, we define the* dual *of h, denoted $\overline{h}$, to be the sequence obtained by replacing every percept (o, r) in h by $(o, -r)$ (in other words: replacing every reward r in h by $-r$).*
2. *Suppose π is an agent. We define the* dual *of π, denoted $\overline{\pi}$, as follows: for each $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$, for each action $y \in \mathcal{A}$, $\overline{\pi}(y|h) = \pi(y|\overline{h})$.*

In plain language, $\overline{\pi}$ acts the way π would act if π wanted to seek punishments and avoid rewards.

Lemma 16. *If x is any agent or history, then $\overline{\overline{x}} = x$.*

Proof. Trivial as $-(-r) = r$ for all real r. ☐

Lemma 17. *For any agent π and any $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$, $P^\pi(\overline{h}) = P^{\overline{\pi}}(h)$.*

Proof. By induction on h. ☐

Definition 18. *An agent π is* self-dual *if $\overline{\pi} = \pi$.*

In plain language, self-dual agents only seek to extremize total reward, without caring about the *sign* of that total reward: a self-dual agent's actions do not change if rewards and punishments are swapped. In particular, any reward-ignoring agent is self-dual. It seems reasonable to expect that an intelligence measure should assign intelligence 0 to reward-ignoring agents. This motivates us to consider intelligence-measure symmetries with respect to duality. The main result in this section will be the equivalence of two such symmetry conditions. But first, we will use mixture agents to characterize self-dual agents (up to equivalence modulo a natural equivalence relation).

Definition 19. *If π and ρ are agents, we say $\pi \equiv \rho$ if the following conditions hold:*

1. *For all $h \in \mathcal{H}$, $P^\pi(h) = 0$ iff $P^\rho(h) = 0$.*
2. *For all $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$, if $P^\pi(h) \neq 0$ then for all $y \in \mathcal{A}$, $\pi(y|h) = \rho(y|h)$.*

Remark 20. *Intuitively, Definition 19 says that $\pi \equiv \rho$ iff the histories which are "possible" for π (in the sense of Remark 2) are exactly the histories which are "possible" for ρ, and $\pi = \rho$ on those histories (π and ρ may differ on "impossible" histories).*

Lemma 21. *$\equiv$ (Definition 19) is an equivalence relation.*

Proof. Straightforward. ☐

Lemma 22. *Let $\vec{w} = (w_1, \ldots, w_n)$ be positive reals, $w_1 + \cdots + w_n = 1$. For every agent π, $\pi \equiv \vec{w}\cdot(\pi, \ldots, \pi)$ (where $(\pi, \ldots, \pi)$ has length n).*

Proof. See Supplementary Materials. ☐

Lemma 23. *For any agent π, $(\frac{1}{2}, \frac{1}{2}) \cdot (\pi, \overline{\pi})$ is self-dual.*

Proof. See Supplementary Materials. ☐

Proposition 24. *(Characterization of self-dual agents modulo $\equiv$) For any agent π, the following are equivalent:*

1. *$\pi \equiv \rho$ for some self-dual agent ρ.*
2. *$\pi \equiv (\frac{1}{2}, \frac{1}{2}) \cdot (\rho, \overline{\rho})$ for some agent ρ.*

Proof. ($\Rightarrow$) Assume $\pi \equiv \rho$ for some self-dual agent ρ. By Lemma 22 $\rho \equiv (\frac{1}{2}, \frac{1}{2}) \cdot (\rho, \rho)$, but $\rho = \overline{\rho}$ by self-duality, so $\rho \equiv (\frac{1}{2}, \frac{1}{2}) \cdot (\rho, \overline{\rho})$. By transitivity of $\equiv$ (Lemma 21), $\pi \equiv (\frac{1}{2}, \frac{1}{2}) \cdot (\rho, \overline{\rho})$.

($\Leftarrow$) Assume $\pi \equiv (\frac{1}{2}, \frac{1}{2}) \cdot (\rho, \overline{\rho})$ for some agent ρ. By Lemma 23, $(\frac{1}{2}, \frac{1}{2}) \cdot (\rho, \overline{\rho})$ is self-dual. ☐

Proposition 24 is analogous to the fact that a function $f : \mathbb{R} \to \mathbb{R}$ is even (i.e. satisfies $f(x) = f(-x)$) iff $f(x) = \frac{1}{2}(g(x) + g(-x))$ for some $g : \mathbb{R} \to \mathbb{R}$.

We will show (Theorem 26) that the following two symmetry conditions are equivalent (improving an informal result of Alexander and Hutter (2021)).

Definition 25. *(Weighted intelligence measure symmetry properties) Let Υ be a weighted intelligence measure.*

1. *Υ is* weakly symmetric *if $\Upsilon(\pi) = 0$ for every self-dual agent π.*
2. *Υ is* strongly symmetric *if $\Upsilon(\overline{\pi}) = -\Upsilon(\pi)$ for every agent π.*

Theorem 26. *A weighted intelligence measure Υ is weakly symmetric iff it is strongly symmetric.*

Proof. (Weak $\Rightarrow$ Strong) Assume Υ is weakly symmetric. Let π be any agent. By Lemma 23, $(\frac{1}{2}, \frac{1}{2}) \cdot (\pi, \overline{\pi})$ is self-dual. So by weak symmetry, $\Upsilon((\frac{1}{2}, \frac{1}{2}) \cdot (\pi, \overline{\pi})) = 0$. Thus by Theorem 14 (part 5),

$$(\tfrac{1}{2}, \tfrac{1}{2}) \cdot \Upsilon((\pi, \overline{\pi})) = \tfrac{1}{2}\Upsilon(\pi) + \tfrac{1}{2}\Upsilon(\overline{\pi}) = 0.$$

So $\Upsilon(\overline{\pi}) = -\Upsilon(\pi)$. By arbitrariness of π, Υ is strongly symmetric.

(Strong $\Rightarrow$ Weak) Trivial. □

The canonical weighted intelligence measure is the universal intelligence measure Υ_U of Legg and Hutter (2007), where each computable environment μ is given weight $2^{-K(\mu)}$, where $K(\mu)$ is μ's Kolmogorov complexity. This depends non-trivially on the background UTM U, prompting Leike and Hutter (2015) to ask: "What are [...] desirable properties of a UTM?" UTMs formalize programming languages, so the question is equivalent to "What are desirable properties of a programming language?" (i.e., *inherently* desirable properties, as opposed to subjectively desirable properties like whether or not white-space matters). Intrinsically desirable UTM properties are elusive; attempts—such as (Müller, 2010)—to find them confirm the difficulty thereof. In the RL context, symmetry conditions on Υ_U are candidate desirable properties for the UTM: U is weakly (resp. strongly) symmetric iff Υ_U is. The equivalence of weak and strong symmetry provides some justification for considering this UTM property to be inherently desirable (in the RL context).

5 DISCERNABILITY AND SEPARABILITY

In this section, we give another application of mixture agents. We define natural notions of *discernability* and *separability* for sets of agents, and we give an interesting characterization of separability in terms of discernability and mixtures. Essentially, we import notions from convex geometry into reinforcement learning. Below, if Π is a set of agents, then Π^c is the set of agents ρ such that $\rho \notin \Pi$.

Definition 27. *A set Π of agents is* discernable *if there exists an environment μ such that for all agents π, ρ:*

1. *V^π_μ and V^ρ_μ are defined.*
2. *If $\pi \in \Pi$ and $\rho \in \Pi^c$, then $V^\pi_\mu \neq V^\rho_\mu$.*

Intuitively, Π is discernable if there is some environment in which no member of Π has the same expected performance as any member of Π^c.

For the next definition, recall that a subset I of the reals is *convex* if the following requirement holds: for all real $i_1 < i_2 < i_3$, if $i_1 \in I$ and $i_3 \in I$, then $i_2 \in I$. Two sets are *disjoint* if they have no point in common.

Definition 28. *A set Π of agents is* separable *if there exists an environment μ and disjoint convex sets I and J of reals such that for every agent π:*

1. *V^π_μ is defined.*
2. *If $\pi \in \Pi$ then $V^\pi_\mu \in I$.*
3. *If $\pi \in \Pi^c$ then $V^\pi_\mu \in J$.*

Equivalently, Π is separable if either: there is some environment where every member of Π outperforms every member of Π^c; or there is some environment where every member of Π underperforms every member of Π^c. Clearly separability implies discernability, but what about the converse? We will use mixtures to state a partial converse.

Definition 29. *A set Π is* closed under mixtures *if the following holds: for all $\vec{w} = (w_1, \ldots, w_n)$ with $w_1 + \cdots + w_n = 1$, for all agents $\vec{\pi} = (\pi_1, \ldots, \pi_n)$, if $\pi_i \in \Pi$ for every $i = 1, \ldots, n$, then $\vec{w} \cdot \vec{\pi} \in \Pi$.*

If we think of agents as points in space, Definition 29 is a universal mixture agent analog of the higher-dimensional convexity notion from convex geometry.

Lemma 30. *For every environment μ and set Π of agents, $S_{\Pi,\mu} = \{r \in \mathbb{R} : \exists \pi_1, \pi_2 \in \Pi \text{ s.t. } V^{\pi_1}_\mu \leq r \leq V^{\pi_2}_\mu\}$ is convex.*

Proof. See Supplementary Material. □

Theorem 31. *(Characterization of Separability) For any set Π of agents, the following are equivalent:*

1. *Π is separable.*
2. *Π is discernable, and both Π and Π^c are closed under mixtures.*

Proof. ($1 \Rightarrow 2$) Assume Π is separable, and let μ, I, J be as in Definition 28, so I and J are disjoint.

To see Π is discernable, let π, ρ be any agents. By condition 1 of Definition 28, V^π_μ and V^ρ_μ are defined. And if $\pi \in \Pi$, $\rho \in \Pi^c$, then by conditions 2 and 3 of Definition 28, $\pi \in I$ and $\rho \in J$. Since I and J are disjoint, $V^\pi_\mu \neq V^\rho_\mu$.

To see Π is closed under mixtures, let $\vec{w} = (w_1, \ldots, w_n)$ and $\vec{\pi} = (\pi_1, \ldots, \pi_n)$ be as in Definition 29 and assume each $\pi_i \in \Pi$. By choice of I, each $V^{\pi_i}_\mu \in I$. By Theorem 14, $V^{\vec{w}\cdot\vec{\pi}}_\mu = \vec{w} \cdot V^{\vec{\pi}}_\mu$. Thus $V^{\vec{w}\cdot\vec{\pi}}_\mu$ is a convex combination of $V^{\pi_1}_\mu, \ldots, V^{\pi_n}_\mu$, which are elements of I. Since I is convex, it follows that $V^{\vec{w}\cdot\vec{\pi}}_\mu \in I$. Since I and J are disjoint, $V^{\vec{w}\cdot\vec{\pi}}_\mu \notin J$, so by choice of J, $\vec{w}\cdot\vec{\pi} \in \Pi$, as desired. A similar argument shows Π^c is closed under mixtures.

($2 \Rightarrow 1$) Assume Π is discernable and both Π and Π^c are closed under mixtures. Since Π is discernable, there is some environment μ as in Definition 27. Let $I = S_{\Pi,\mu}$, $J = S_{\Pi^c,\mu}$ as in Lemma 30, so I and J are convex. From the definition of I and J, clearly $V^\pi_\mu \in I$ for all $\pi \in \Pi$ and $V^\rho_\mu \in J$ for all $\pi \in \Pi^c$. It only remains to show I and J are disjoint. Assume not. Then there is some $r \in I \cap J$. By definition of I, there are $\pi_1, \pi_2 \in \Pi$ such that $V^{\pi_1}_\mu \leq r \leq V^{\pi_2}_\mu$, and by definition of J, there are $\rho_1, \rho_2 \in \Pi^c$ such that $V^{\rho_1}_\mu \leq r \leq V^{\rho_2}_\mu$. By basic algebra, there is a real $\alpha \in [0,1]$ such that $\alpha V^{\pi_1}_\mu + (1-\alpha)V^{\pi_2}_\mu = r$. Let $\pi = (\alpha, 1-\alpha)\cdot(\pi_1, \pi_2)$. By Theorem 14 (part 4), $V^\pi_\mu = \alpha V^{\pi_1}_\mu + (1-\alpha)V^{\pi_2}_\mu = r$. And since Π is closed under mixtures, $\pi \in \Pi$. By identical reasoning using $V^{\rho_1}_\mu \leq r \leq V^{\rho_2}_\mu$, there exists some $\rho \in \Pi^c$ such that $V^\rho_\mu = r$. But since μ satisfies condition 2 of Definition 27, and $\pi \in \Pi$ and $\rho \in \Pi^c$, this forces $V^\pi_\mu \neq V^\rho_\mu$, absurd. □

6 LOCAL EXTREMA AND LATTICE POINTS

Definition 32. *If π is an agent, $h_0 \in (\mathcal{EA})^*\mathcal{E}$, and m is a probability distribution on $\mathcal{A}$, we write $\pi^{h_0 \mapsto m}$ for the function which is identical to π except that m decides the action distribution for h_0, that is,*

$$\pi^{h_0\mapsto m}(y|h) = \begin{cases} \pi(y|h) & \text{if } h \neq h_0, \\ m(y) & \text{if } h = h_0. \end{cases}$$

In plain language, $\pi^{h_0\mapsto m}$ is the result of changing π's output $\pi(\cdot|h_0)$ to m, but otherwise leaving π unchanged.

Lemma 33. *$\pi^{h_0\mapsto m}$ (as in Definition 32) is an agent.*

Proof. Trivial. □

Definition 34. *Suppose $\vec{m} = (m_1, \ldots, m_n)$ are probability distributions on $\mathcal{A}$ and $\vec{w} = (w_1, \ldots, w_n)$ are positive reals with $w_1 + \cdots + w_n = 1$. By $\vec{w}\cdot\vec{m}$ we mean the function on $\mathcal{A}$ defined by*

$$(\vec{w}\cdot\vec{m})(y) = w_1 m_1(y) + \cdots + w_n m_n(y).$$

Lemma 35. *If $\vec{m}$, $\vec{w}$ are as in Definition 34 then $\vec{w}\cdot\vec{m}$ is a probability distribution on $\mathcal{A}$.*

Proof. See Supplementary Materials. □

Definition 36. *For any agent π, for any $h \in (\mathcal{EA})^*\mathcal{E}$, for any probability distributions $\vec{m} = (m_1, \ldots, m_n)$ on $\mathcal{A}$, let $\pi^{h\mapsto\vec{m}} = (\pi^{h\mapsto m_1}, \ldots, \pi^{h\mapsto m_n})$.*

The following proposition shows that for any particular history h and agent π, for any decomposition of $\pi(\cdot|h)$ into a weighted sum of probability distributions $m_1, \ldots, m_n$, π has the same intelligence as the weighted mixture of the corresponding n agents $\pi^{h\mapsto\vec{m}}$.

Proposition 37. *Let Υ be any weighted intelligence measure, let π be any agent, and let $h \in (\mathcal{EA})^*\mathcal{E}$. Suppose $\vec{m}$ and $\vec{w}$ are as in Definition 34. If $\vec{w}\cdot\vec{m} = \pi(\cdot|h)$, then $\Upsilon(\pi) = \Upsilon(\vec{w}\cdot\pi^{h\mapsto\vec{m}})$.*

Proof. See Supplementary Materials. □

Definition 38. *Suppose π and $\pi_1, \ldots, \pi_n$ are agents and Υ is a weighted intelligence measure. We say $\pi \succ_\Upsilon \pi_1, \ldots, \pi_n$ if both:*

1. *$\Upsilon(\pi) \geq \Upsilon(\pi_i)$ for each $i = 1, \ldots, n$; and*
2. *$\Upsilon(\pi) > \Upsilon(\pi_i)$ for some $i = 1, \ldots, n$.*

We define $\pi \prec_\Upsilon \pi_1, \ldots, \pi_n$ likewise (change $\geq$/$>$ to $\leq$/$<$).

Proposition 39. *Let Υ be any weighted intelligence measure and let π be an agent. Let $h \in (\mathcal{EA})^*\mathcal{E}$. For any probability distributions $\vec{m} = (m_1, \ldots, m_n)$ on $\mathcal{A}$, for any positive reals $\vec{w} = (w_1, \ldots, w_n)$ with $w_1 + \cdots + w_n = 1$, if $\vec{w}\cdot\vec{m} = \pi(\cdot|h)$, then $\pi \not\succ_\Upsilon \pi^{h\mapsto m_1}, \ldots, \pi^{h\mapsto m_n}$ and $\pi \not\prec_\Upsilon \pi^{h\mapsto m_1}, \ldots, \pi^{h\mapsto m_n}$.*

Proof. If $\pi \succ_\Upsilon \pi^{h\mapsto m_1}, \ldots, \pi^{h\mapsto m_n}$ then this implies

$$\begin{aligned} \Upsilon(\pi) &= \Upsilon(\vec{w}\cdot\pi^{h\mapsto\vec{m}}) && \text{(Proposition 37)} \\ &= \vec{w}\cdot\Upsilon(\pi^{h\mapsto\vec{m}}) && \text{(Theorem 14)} \\ &< w_1\Upsilon(\pi) + \cdots + w_n\Upsilon(\pi) && \text{(Assumption)} \\ &= \Upsilon(\pi), && (w_1 + \cdots + w_n = 1) \end{aligned}$$

absurd. Similar reasoning holds for $\prec_\Upsilon$. □

Definition 40. *(Local intelligence extrema)*

1. *We make the space of all agents into a metric space by defining the distance from agent π to agent ρ to be $d(\pi, \rho) = \sup_{h\in(\mathcal{EA})^*\mathcal{E}, y\in\mathcal{A}} |\pi(y|h) - \rho(y|h)|$.*
2. *Suppose Υ is a weighted intelligence measure. An agent π is a* strict local maximum *(resp.* strict local minimum*) of Υ if there is some real $\epsilon > 0$ such that for every agent $\rho \not\equiv \pi$ (recall Definition 19), if*

$d(\rho, \pi) < \epsilon$ then $\Upsilon(\pi) > \Upsilon(\rho)$ (resp. $\Upsilon(\pi) < \Upsilon(\rho)$). If π is a strict local maximum or minimum of Υ then π is a strict local extremum *of Υ.*

Definition 41. *An agent π is* deterministic in all possible histories *if the following condition holds. For all $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$, if $P^\pi(h) \neq 0$ then for all $y \in \mathcal{A}$, $\pi(y|h) \in \{0, 1\}$.*

Note that an agent π can be deterministic in all possible histories (Definition 41) and still assign a probability $0 < \pi(y|h) < 1$, provided $P^\pi(h) = 0$ (recall Remark 2). It is easy to show that π is deterministic in all possible histories iff $\pi \equiv \rho$ for some strictly deterministic ρ (i.e., some ρ such that $\rho(y|h) \in \{0, 1\}$ for all y, h).

Theorem 42 below sheds light on the geometry of RL agent intelligence. By considering agent π's coordinates to be the values $\pi(y|h)$ for all y, h, we can view agents as inhabiting infinite-dimensional Euclidean space. An agent π is a *lattice point* (i.e., a point with integer coordinates) iff π is strictly deterministic. We can picture $z = \Upsilon(\pi)$ as a "surface" above agent-space (more precisely: a surface above agent-space in some places, below it in others, and which intersects it in the "z-intercept" $\Upsilon(\pi) = 0$). Theorem 42 says that, modulo $\equiv$, $z = \Upsilon(\pi)$ cannot have any "hyperridges" or "hypertroughs" above non-lattice points.

Theorem 42. *("Strict local extrema are deterministic") For any weighted intelligence measure Υ, for any agent π, if π is a strict local extremum of Υ, then π is deterministic in all possible histories.*

Proof. Assume π is not deterministic in all possible histories, so there exist $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$ and $y_0 \in \mathcal{A}$ such that $P^\pi(h) \neq 0$ and $0 < \pi(y_0|h) < 1$. We will show π is not a strict local maximum (a similar argument shows π is not a strict local minimum) of Υ. Since $0 < \pi(y_0|h) < 1$ and $\pi(\cdot|h) \in \Delta\mathcal{A}$, there must be some $y_1 \in \mathcal{A}$, $y_1 \neq y_0$, such that $0 < \pi(y_1|h) < 1$. Let $\epsilon > 0$. Since $0 < \pi(y_0|h) < 1$ and $0 < \pi(y_1|h) < 1$, it follows that there is some $0 < \epsilon' \leq \epsilon$ such that $0 < \pi(y_0|h) \pm \epsilon' < 1$ and $0 < \pi(y_1|h) \pm \epsilon' < 1$. Define $m_1, m_2 : \mathcal{A} \to \mathbb{R}$ by

$$m_i(y) = \begin{cases} \pi(y|h) + (-1)^i\epsilon' & \text{if } y = y_0, \\ \pi(y|h) - (-1)^i\epsilon' & \text{if } y = y_1, \\ \pi(y|h) & \text{otherwise.} \end{cases}$$

By choice of ϵ' it follows that $m_1, m_2 \in \Delta\mathcal{A}$. Let $\vec{w} = (\frac{1}{2}, \frac{1}{2})$, $\vec{m} = (m_1, m_2)$; clearly $\vec{w} \cdot \vec{m} = \pi(\cdot|h)$. By Prop. 39, $\pi \not\succ_\Upsilon \pi^{h \mapsto m_1}, \pi^{h \mapsto m_2}$, thus $\Upsilon(\pi) \leq \Upsilon(\pi^{h \mapsto m_i})$ for some $i \in \{1, 2\}$. Clearly $\pi \not\equiv \pi^{h \mapsto m_i}$ and $d(\pi, \pi^{h \mapsto m_i}) = \epsilon' \leq \epsilon$. By arbitrariness of ϵ, π is not a strict local maximum of Υ. □

In a sense, the above proof is a lack-of-emergent-behavior proof. The non-determinacy of π allows us to perturb π very slightly in two opposing directions, in such a way that π is the weighted mixture of the two perturbations. If, say, both perturbations strictly reduced π's intelligence, then their mixture would exhibit emergent behavior (namely: "behave at least as intelligently as π") of a type ruled out by Theorem 14.

We have worked in this section using the metric of Definition 40 for simplicity. Similar reasoning would apply to various other metrics as well.

7 UNIVERSAL MIXTURE-ENVIRONMENTS

Definition 43. *An environment μ is* strongly well-behaved *if μ is well-behaved and for all agents π and all $t \in \mathbb{N}$, $-1 \leq V^\pi_{\mu,t} \leq 1$. A weighted intelligence measure Υ is* strongly well-behaved *if corresponding weights $\{w_\mu\}_{\mu \in W}$ (as in Definition 8) exist such that $\sum_{\mu \in W} w_\mu = 1$ and such that $w_\mu = 0$ for all μ not strongly well-behaved (informally: the weights underlying Υ sum to 1, and any environment not strongly well-behaved has weight 0).*

If μ never gives negative rewards, then $-1 \leq V^\pi_{\mu,t} \leq 1$ is equivalent to $-1 \leq V^\pi_\mu \leq 1$. Thus if the reward-space $\mathcal{R}$ is $\subseteq [0, 1]$ (as in (Legg and Hutter, 2007)), then every nonzero weighted intelligence measure is a constant multiple of a strongly well-behaved one. We will show (Theorem 49) that for every strongly well-behaved Υ, there is an environment μ_Υ such that for all agents π, $\Upsilon(\pi) = V^\pi_{\mu_\Upsilon}$.

In this section, let $\mathscr{W}$ be the set infinite sequences $\vec{w} = (w_1, w_2, \ldots)$ with each $w_i > 0$ real and $\sum_{i=1}^\infty w_i = 1$. For any $w \in \mathscr{W}$ and any bounded sequence $\vec{v} = (v_1, v_2, \ldots)$, we define the dot product $\vec{w} \cdot \vec{v} = \sum_{i=1}^\infty w_i v_i$ (the boundedness of $\vec{v}$ implies this sum converges).

Definition 44. *(Compare Definition 9) Let $\vec{w} \in \mathscr{W}$, π an agent, μ an environment, $h \in \mathcal{H}$, $t \in \mathbb{N}$, and $\vec{\mu} = (\mu_1, \mu_2, \ldots)$ an infinite sequence of environments. Define:*

- $P_{\vec{\mu}}(h) = (P_{\mu_1}(h), P_{\mu_2}(h), \ldots)$.
- $P^\pi_{\vec{\mu}}(h) = (P^\pi_{\mu_1}(h), P^\pi_{\mu_2}(h), \ldots)$.
- $V^\pi_{\vec{\mu},t} = (V^\pi_{\mu_1,t}, V^\pi_{\mu_2,t}, \ldots)$.
- $V^\pi_{\vec{\mu}} = (V^\pi_{\mu_1}, V^\pi_{\mu_2}, \ldots)$ *if every $V^\pi_{\mu_i}$ is defined.*

Definition 45. *(Mixture environments—compare Definition 11) Assume $\vec{w} \in \mathscr{W}$ and $\vec{\mu} = (\mu_1, \mu_2, \ldots)$ is an infinite sequence of environments. Define an environment $\vec{w} \cdot \vec{\mu}$ by:*

$$(\vec{w} \cdot \vec{\mu})(x|h) = \begin{cases} \dfrac{\vec{w} \cdot P_{\vec{\mu}}(hx)}{\vec{w} \cdot P_{\vec{\mu}}(h)} & \textit{if } \vec{w} \cdot P_{\vec{\mu}}(h) \neq 0, \\ 1/|\mathcal{E}| & \textit{otherwise.} \end{cases}$$

Lemma 46. *(Compare Lemma 13) $\vec{w} \cdot \vec{\mu}$ (as in Definition 45) is indeed an environment.*

Proof. See Supplementary Materials. □

To prove Lemma 48 below, we will use Tannery's Theorem, a result from real analysis (Bromwich, 2005).

Lemma 47. *(Tannery's Theorem) Let $\{a_i : \mathbb{N} \to \mathbb{R}\}_{i=1}^{\infty}$ be a sequence of sequences such each $\lim_{t\to\infty} a_i(t)$ converges. Assume $\{w_i\}_{i=1}^{\infty}$ satisfies $\sum_{i=1}^{\infty} w_k < \infty$ and for all $i > 0$, for all $t \in \mathbb{N}$, $|a_i(t)| \leq w_k$. Then*

$$\lim_{t\to\infty} \sum_{i=1}^{\infty} a_i(t) = \sum_{i=1}^{\infty} \lim_{t\to\infty} a_i(t).$$

Lemma 48. *(Compare Theorem 14) Let $\vec{w} \in \mathscr{W}$, let $\vec{\mu} = (\mu_1, \mu_2, \ldots)$ be a sequence of strongly well-behaved environments, and let π be any agent. Then:*

1. *For all $h \in \mathcal{H}$, $P_{\vec{w}\cdot\vec{\mu}}(h) = \vec{w} \cdot P_{\vec{\mu}}(h)$.*
2. *For all $h \in \mathcal{H}$, $P^{\pi}_{\vec{w}\cdot\vec{\mu}}(h) = \vec{w} \cdot P^{\pi}_{\vec{\mu}}(h)$.*
3. *For all $t \in \mathbb{N}$, $V^{\pi}_{\vec{w}\cdot\vec{\mu},t} = \vec{w} \cdot V^{\pi}_{\vec{\mu},t}$.*
4. *$V^{\pi}_{\vec{w}\cdot\vec{\mu}} = \vec{w} \cdot V^{\pi}_{\vec{\mu}}$.*

Proof. For (1)–(3), see Supplementary Materials. For (4):

$$\begin{aligned}
V^{\pi}_{\vec{w}\cdot\vec{\mu}} &= \lim_{t\to\infty} V^{\pi}_{\vec{w}\cdot\vec{\mu},t} && \text{(Definition 6)}\\
&= \lim_{t\to\infty} \vec{w} \cdot V^{\pi}_{\vec{\mu},t} && \text{(By (3))}\\
&= \lim_{t\to\infty} \sum_{i=1}^{\infty} w_i V^{\pi}_{\mu_i,t}. && \text{(Definition 44)}
\end{aligned}$$

For each $i \geq 1$, define $a_i : \mathbb{N} \to \mathbb{R}$ by $a_i(t) = w_i V^{\pi}_{\mu_i,t}$. Since μ_i is strongly well-behaved, each $-1 \leq V^{\pi}_{\mu_i,t} \leq 1$. It follows that for all $t \in \mathbb{N}$, $|a_i(t)| \leq w_i$. Furthermore, $\sum_{i=1}^{\infty} w_i = 1 < \infty$ by Definition of $\mathscr{W}$. Thus

$$\begin{aligned}
&\lim_{t\to\infty} \sum_{i=1}^{\infty} w_i V^{\pi}_{\mu_i,t}\\
&= \sum_{i=1}^{\infty} \lim_{t\to\infty} w_i V^{\pi}_{\mu_i,t} && \text{(Tannery's Theorem)}\\
&= \sum_{i=1}^{\infty} w_i \lim_{t\to\infty} V^{\pi}_{\mu_i,t} && \text{(Algebra)}\\
&= \sum_{i=1}^{\infty} w_i V^{\pi}_{\mu_i} && \text{(Definition 6)}\\
&= \vec{w} \cdot V^{\pi}_{\vec{\mu}}. && \text{(Definition 44)}
\end{aligned}$$

So $V^{\pi}_{\vec{w}\cdot\vec{\mu}} = \vec{w} \cdot V^{\pi}_{\vec{\mu}}$. □

Just as Theorem 14 (parts 4–5) shows that Definition 11 provides a way to mix agents without emergent behavior, in the same way, Lemma 48 shows that Definition 45 provides a way to mix environments without emergent behavior.

Theorem 49. *For any strongly well-behaved weighted intelligence measure Υ, there is an environment μ_Υ such that for every agent π, $\Upsilon(\pi) = V^{\pi}_{\mu_\Upsilon}$.*

Proof. Let $\vec{\mu} = (\mu_1, \mu_2, \ldots)$ enumerate all strongly well-behaved environments (a countable set since every well-behaved environment is Turing computable). Let $(w_\mu)_{\mu\in W}$ be as in Definition 43. Let $\vec{w} = (w_{\mu_1}, w_{\mu_2}, \ldots)$, thus $\vec{w} \in \mathscr{W}$. Then for any agent π,

$$\begin{aligned}
V^{\pi}_{\vec{w}\cdot\vec{\mu}} &= \vec{w} \cdot V^{\pi}_{\vec{\mu}} && \text{(Lemma 48)}\\
&= \sum_{i=1}^{\infty} w_i V^{\pi}_{\mu_i} && \text{(Definition 44)}\\
&= \Upsilon(\pi), && \text{(Definition 43)}
\end{aligned}$$

so $\mu_\Upsilon = \vec{w} \cdot \vec{\mu}$ works. □

8 SUMMARY

We introduced (Definition 11) an operation which takes a finite sequence of RL agents and a finite sequence of weights, and which outputs a new agent, which can be thought of as a weighted mixture agent, with the property that in any environment, "The expected reward of a weighted mixture is the weighted average of the expected rewards" (Theorem 14 part 4). Thus if intelligence is measured in terms of performance, "The intelligence of a weighted mixture is the weighted average of the intelligences" (Theorem 14 part 5). This construction enabled us to prove a number of results about the geometry of RL agent intelligence measures, namely, results about intelligence symmetry (Theorem 26), convexity (Theorem 31), and strict local extrema (Theorem 42). Finally, by applying the same mixture idea to environments instead of agents, we established (Theorem 49) that for a large class of performance-based intelligence measures, there exist universal mixture environments, i.e., environments in which every agent's total expected reward in fact equals the agent's intelligence according to said measure.

Acknowledgements

This work has been supported in parts by ARC grant DP150104590.

References

Alexander, S. A. and Hutter, M. (2021). Reward-punishment symmetric universal intelligence. In *AGI*.

Bouricius, T. (2013). Democracy through multi-body sortition: Athenian lessons for the modern day. *Journal of Deliberative Democracy*, 9:11.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W. (2016). OpenAI gym. *Preprint*.

Bromwich, T. J. I. (2005). *An introduction to the theory of infinite series*, volume 335. American Mathematical Society.

Chen, M., Shlezinger, N., Poor, H. V., Eldar, Y. C., and Cui, S. (2021). Communication-efficient federated learning. *Proceedings of the National Academy of Sciences*, 118.

Dayan, P. and Hinton, G. E. (1992). Feudal reinforcement learning. In *NIPS*.

Flanigan, B., Gölz, P., Gupta, A., Hennig, B., and Procaccia, A. (2021). Fair algorithms for selecting citizens' assemblies. *Nature*, 596.

Hansen, M. H. (1991). *The Athenian democracy in the age of Demosthenes : structure, principles, and ideology / Mogens Herman Hansen ; translated by J.A. Crook*. B. Blackwell Oxford, UK ; Cambridge, USA.

Hernández-Orallo, J., Dowe, D. L., Espana-Cubillo, S., Hernández-Lloreda, M. V., and Insa-Cabrera, J. (2011). On more realistic environment distributions for defining, evaluating and developing intelligence. In *AGI*.

Hutter, M. (2003). On the existence and convergence of computable universal priors. In Gavaldà, R., Jantke, K. P., and Takimoto, E., editors, *Algorithmic Learning Theory, 14th International Conference, ALT 2003, Sapporo, Japan, October 17-19, 2003, Proceedings*, volume 2842 of *Lecture Notes in Computer Science*, pages 298–312. Springer.

Hutter, M. (2009). Discrete MDL predicts in total variation. *Advances in Neural Information Processing Systems*, 22.

Johnson, F. and Dana, K. J. (2020). Feudal steering: Hierarchical learning for steering angle prediction. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4316–4325.

Kok, J. R. and Vlassis, N. A. (2006). Collaborative multiagent reinforcement learning by payoff propagation. *J. Mach. Learn. Res.*, 7:1789–1828.

Lanctot, M., Zambaldi, V. F., Gruslys, A., Lazaridou, A., Tuyls, K., Pérolat, J., Silver, D., and Graepel, T. (2017). A unified game-theoretic approach to multiagent reinforcement learning. In *NIPS*.

Legg, S. and Hutter, M. (2005). A universal measure of intelligence for artificial agents. In Kaelbling, L. P. and Saffiotti, A., editors, *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005*, pages 1509–1510. Professional Book Center.

Legg, S. and Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and machines*, 17(4):391–444.

Leike, J. and Hutter, M. (2015). Bad universal priors and notions of optimality. In *Conference on Learning Theory*, pages 1244–1259. PMLR.

Li, M. and Vitányi, P. (2008). *An introduction to Kolmogorov complexity and its applications*. Springer.

Li, Z., Shi, L., Cristea, A. I., and Zhou, Y. (2021). A survey of collaborative reinforcement learning: Interactive methods and design patterns. In *Designing Interactive Systems Conference 2021*, DIS '21, page 1579–1590, New York, NY, USA. Association for Computing Machinery.

Littman, M. L. (1994). Markov games as a framework for multi-agent reinforcement learning. In *ICML*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., Graves, A., Riedmiller, M., Fidjeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518:529–33.

Müller, M. (2010). Stationary algorithmic probability. *Theoretical Computer Science*, 411(1):113–130.

Qiu, W., Wang, X., Yu, R., He, X., Wang, R., An, B., Obraztsova, S., and Rabinovich, Z. (2021). Rmix: Learning risk-sensitive policies for cooperative reinforcement learning agents. In *NeurIPS*.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Driessche, G., Graepel, T., and Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, 550:354–359.

Sintomer, Y. (2018). From deliberative to radical democracy? sortition and politics in the twenty-first century. *Politics & Society*, 46:337 – 357.

Vinyals, O., Babuschkin, I., Czarnecki, W., Mathieu, M., Dudzik, A., Chung, J., Choi, D., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J., Jaderberg, M., and Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575.

Weiss, G. (1993). Collective learning of action sequences. *[1993] Proceedings. The 13th International Conference on Distributed Computing Systems*, pages 203–209.

Zhang, K., Yang, Z., and Başar, T. (2021). *Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms*, pages 321–384. Springer International Publishing, Cham.

A SUPPLEMENTARY MATERIAL

Here, we present detailed proofs missing from the main text due to length limit.

A.1 Proof of Lemma 5

Proof. By induction on h.

Case 1: $h=\varepsilon$. Then the lemma is trivial.

Case 2: $h=gx$ for some $x\in\mathcal{E}$. Then

$$\begin{aligned}
P^\pi_\mu(h) &= P^\pi_\mu(g)\mu(x|g) && \text{(Definition 4)}\\
&= P^\pi(g)P_\mu(g)\mu(x|g) && \text{(Induction)}\\
&= P^\pi(h)P_\mu(g)\mu(x|g) && \text{(Definition 4)}\\
&= P^\pi(h)P_\mu(h). && \text{(Definition 4)}
\end{aligned}$$

Case 3: $h=gy$ for some $y\in\mathcal{A}$. Similar to Case 2. □

A.2 Proof of Lemma 13

Proof. Let $h\in(\mathcal{E}\mathcal{A})^*\mathcal{E}$. Clearly $(\vec{w}\cdot\vec{\pi})(y|h)\geq 0$ for all $y\in\mathcal{A}$. It remains to show $\sum_{y\in\mathcal{A}}(\vec{w}\cdot\vec{\pi})(y|h)=1$.

Case 1: $\vec{w}\cdot P^{\vec{\pi}}(h)=0$. Then each $(\vec{w}\cdot\vec{\pi})(y|h)=1/|\mathcal{A}|$ so the claim is immediate.

Case 2: $\vec{w}\cdot P^{\vec{\pi}}(h)\neq 0$. Then

$$\begin{aligned}
&\sum_{y\in\mathcal{A}}(\vec{w}\cdot\vec{\pi})(y|h)\\
&= \sum_{y\in\mathcal{A}}\frac{\vec{w}\cdot P^{\vec{\pi}}(hy)}{\vec{w}\cdot P^{\vec{\pi}}(h)} && \text{(Definition 11)}\\
&= \sum_{y\in\mathcal{A}}\frac{\vec{w}\cdot (P^{\pi_1}(hy),\ldots,P^{\pi_n}(hy))}{\vec{w}\cdot P^{\vec{\pi}}(h)} && \text{(Definition 9)}\\
&= \sum_{y\in\mathcal{A}}\frac{w_1P^{\pi_1}(h)\pi_1(y|h)+\cdots+w_nP^{\pi_n}(h)\pi_n(y|h)}{\vec{w}\cdot P^{\vec{\pi}}(h)} && \text{(Definition 4)}\\
&= \frac{w_1P^{\pi_1}(h)\left(\sum_{y\in\mathcal{A}}\pi_1(y|h)\right)+\cdots+w_nP^{\pi_n}(h)\left(\sum_{y\in\mathcal{A}}\pi_n(y|h)\right)}{\vec{w}\cdot P^{\vec{\pi}}(h)} && \text{(Algebra)}\\
&= \frac{w_1P^{\pi_1}(h)\cdot 1+\cdots+w_nP^{\pi_n}(h)\cdot 1}{\vec{w}\cdot P^{\vec{\pi}}(h)}=\frac{\vec{w}\cdot P^{\vec{\pi}}(h)}{\vec{w}\cdot P^{\vec{\pi}}(h)}=1. && \text{(π_i are agents)}
\end{aligned}$$

□

A.3 Proof of Lemma 22

Proof. Recall that the real numbers satisfy the so-called *null-factor law*: for all real numbers a and b, if $ab=0$, then $a=0$ or $b=0$. In other words, the product of two nonzero real numbers can never be zero.

Write $\vec{\pi}$ for $(\pi,\ldots,\pi)$. We prove conditions 1 and 2 of Definition 19 simultaneously by induction on h.

Case 1: $h=\varepsilon$. Then $P^\pi(h)=\vec{w}\cdot P^{\vec{\pi}}(h)=1\neq 0$, so vacuously $P^\pi(h)=0$ iff $\vec{w}\cdot P^{\vec{\pi}}(h)=0$ (proving condition 1). For condition 2, there is nothing to check, since $\varepsilon\notin(\mathcal{E}\mathcal{A})^*\mathcal{E}$.

Case 2: $h=h_0y_0$ for some $h_0\in(\mathcal{E}\mathcal{A})^*\mathcal{E}$, $y_0\in\mathcal{A}$. For condition 2, there is nothing to prove, since $h\notin(\mathcal{E}\mathcal{A})^*\mathcal{E}$. For condition 1, we consider two cases.

Subcase 2.1: $P^{\pi}(h_0) = 0$. By induction, condition 1 holds for h_0, so $\vec{w} \cdot P^{\vec{\pi}}(h_0) = 0$. By Definition 4, $P^{\pi}(h) = P^{\pi}(h_0)\pi(y_0|h_0) = 0$ and $P^{\vec{w}\cdot\vec{\pi}}(h) = 0(\vec{w}\cdot\vec{\pi})(y_0|h_0) = 0$. So $P^{\pi}(h) = 0$ iff $P^{\vec{w}\cdot\vec{\pi}}(h) = 0$.

Subcase 2.2: $P^{\pi}(h_0) \neq 0$. Then

$$\begin{aligned}
P^{\vec{w}\cdot\vec{\pi}}(h) &= \vec{w} \cdot P^{\vec{\pi}}(h) && \text{(Theorem 14)}\\
&= w_1P^{\pi}(h) + \cdots + w_nP^{\pi}(h) && \text{(Def. of } \vec{w} \text{ and } \vec{\pi})\\
&= P^{\pi}(h) && (w_1 + \cdots + w_n = 1)\\
&= P^{\pi}(h_0)\pi(y_0|h_0). && \text{(Definition 4)}
\end{aligned}$$

Since $P^{\pi}(h_0) \neq 0$, by the null-factor law, it follows that $P^{\vec{w}\cdot\vec{\pi}}(h) = 0$ iff $P^{\pi}(h) = 0$ iff $\pi(y_0|h_0) = 0$.

Case 3: $h = h_0x$ for some $h_0 \in (\mathcal{E}\mathcal{A})^*$, $x \in \mathcal{E}$. By induction, conditions 1 and 2 hold for h_0. By Definition 4, $P^{\pi}(h) = P^{\pi}(h_0)$ and $P^{\vec{w}\cdot\vec{\pi}}(h) = P^{\vec{w}\cdot\vec{\pi}}(h_0)$, so condition 1 for h follows.

For condition 2, assume $P^{\pi}(h) \neq 0$ and let $y \in \mathcal{A}$. By choice of $\vec{w}$ and $\vec{\pi}$, $\vec{w} \cdot P^{\vec{\pi}}(h) = w_1P^{\pi}(h) + \cdots + w_nP^{\pi}(h) = P^{\pi}(h)$. So, since $P^{\pi}(h) \neq 0$, $\vec{w} \cdot P^{\vec{\pi}}(h) \neq 0$. Thus

$$\begin{aligned}
(\vec{w}\cdot\vec{\pi})(y|h) &= \frac{\vec{w} \cdot P^{\vec{\pi}}(hy)}{\vec{w} \cdot P^{\vec{\pi}}(h)} && \text{(Definition 11)}\\
&= \frac{w_1P^{\pi}(hy) + \cdots + w_nP^{\pi}(hy)}{w_1P^{\pi}(h) + \cdots + w_nP^{\pi}(h)} && \text{(Def. of } \vec{w} \text{ and } \vec{\pi})\\
&= \frac{w_1P^{\pi}(h) + \cdots + w_nP^{\pi}(h)}{w_1P^{\pi}(h) + \cdots + w_nP^{\pi}(h)}\pi(y|h) && \text{(Definition 4)}\\
&= \pi(y|h).
\end{aligned}$$

□

A.4 Proof of Lemma 23

Proof. Let $\vec{w} = (\frac{1}{2}, \frac{1}{2})$. For any $h \in (\mathcal{E}\mathcal{A})^*\mathcal{E}$ and $y \in \mathcal{A}$, we claim

$$\overline{\vec{w}\cdot(\pi,\overline{\pi})}(y|h) = (\vec{w}\cdot(\pi,\overline{\pi}))(y|h).$$

Noting that $\vec{w} \cdot P^{(\pi,\overline{\pi})}(h) = \frac{1}{2}P^{\pi}(h) + \frac{1}{2}P^{\overline{\pi}}(h)$ and $\vec{w} \cdot P^{(\pi,\overline{\pi})}(\overline{h}) = \frac{1}{2}P^{\pi}(\overline{h}) + \frac{1}{2}P^{\overline{\pi}}(\overline{h})$, Lemmas 16 and 17 imply that

$$\vec{w} \cdot P^{(\pi,\overline{\pi})}(h) = \vec{w} \cdot P^{(\pi,\overline{\pi})}(\overline{h}).$$

So if $\vec{w} \cdot P^{(\pi,\overline{\pi})}(\overline{h}) = 0$ then $\vec{w} \cdot P^{(\pi,\overline{\pi})}(h) = 0$ and it follows from Definition 11 and Lemma 17 that $\overline{\vec{w}\cdot(\pi,\overline{\pi})}(y|h) = (\vec{w}\cdot(\pi,\overline{\pi}))(y|h) = 1/|\mathcal{A}|$. So assume $\vec{w} \cdot P^{(\pi,\overline{\pi})}(\overline{h}) \neq 0$. Then:

$$\begin{aligned}
\overline{\vec{w}\cdot(\pi,\overline{\pi})}(y|h) &= (\vec{w}\cdot(\pi,\overline{\pi}))(y|\overline{h}) && \text{(Definition 15)}\\
&= \frac{\frac{1}{2}P^{\pi}(\overline{h}y) + \frac{1}{2}P^{\overline{\pi}}(\overline{h}y)}{\frac{1}{2}P^{\pi}(\overline{h}) + \frac{1}{2}P^{\overline{\pi}}(\overline{h})} && \text{(Definition 11)}\\
&= \frac{\frac{1}{2}P^{\pi}(\overline{hy}) + \frac{1}{2}P^{\overline{\pi}}(\overline{hy})}{\frac{1}{2}P^{\pi}(\overline{h}) + \frac{1}{2}P^{\overline{\pi}}(\overline{h})} && \text{(Clearly } \overline{h}y = \overline{hy})\\
&= \frac{\frac{1}{2}P^{\overline{\pi}}(hy) + \frac{1}{2}P^{\overline{\overline{\pi}}}(hy)}{\frac{1}{2}P^{\overline{\pi}}(h) + \frac{1}{2}P^{\overline{\overline{\pi}}}(h)} && \text{(Lemma 17)}\\
&= (\vec{w}\cdot(\overline{\overline{\pi}},\overline{\pi}))(y|h) && \text{(Definition 11)}\\
&= (\vec{w}\cdot(\pi,\overline{\pi}))(y|h). && \text{(Lemma 16)}
\end{aligned}$$

□

A.5 Proof of Lemma 30

Proof. Assume $i_1 < i_2 < i_3$ are reals with $i_1, i_3 \in S_{\Pi,\mu}$; we must show $i_2 \in S_{\Pi,\mu}$. Since $i_1 \in S_{\Pi,\mu}$, there exist agents $\pi_1, \pi_2 \in \Pi$ such that $V^{\pi_1}_{\mu} \leq i_1 \leq V^{\pi_2}_{\mu}$. And since $i_3 \in S_{\Pi,\mu}$, there exist agents $\rho_1, \rho_2 \in \Pi$ such that $V^{\rho_1}_{\mu} \leq i_3 \leq V^{\rho_2}_{\mu}$. Then $\pi_1, \rho_2 \in \Pi$ satisfy $V^{\pi_1}_{\mu} \leq i_2 \leq V^{\rho_2}_{\mu}$, showing $i_2 \in S_{\Pi,\mu}$ as desired. □

A.6 Proof of Lemma 35

Proof. Clearly for every $y \in \mathcal{A}$, $(\vec{w} \cdot \vec{m})(y) = w_1 m_1(y) + \cdots + w_n m_n(y)$ is a nonnegative real. It remains to show $\sum_{y\in\mathcal{A}}(\vec{w} \cdot \vec{m})(y) = 1$. We compute:

$$\begin{aligned}
&\sum_{y\in\mathcal{A}}(\vec{w} \cdot \vec{m})(y) \\
&= \sum_{y\in\mathcal{A}} w_1 m_1(y) + \cdots + w_n m_n(y) && \text{(Definition 34)} \\
&= w_1 \left(\sum_{y\in\mathcal{A}} m_1(y)\right) + \cdots + w_n \left(\sum_{y\in\mathcal{A}} m_n(y)\right) && \text{(Basic Algebra)} \\
&= w_1 + \cdots + w_n && (m_1, \ldots, m_n \text{ are probability distr's}) \\
&= 1.
\end{aligned}$$

□

A.7 Proof of Proposition 37

The following auxiliary lemmas will be used in our proof of Proposition 37.

Lemma 50. *Suppose π, h_0, m are as in Definition 32. Let $h \in \mathcal{H}$ be such that for every $y \in \mathcal{A}$, $h_0 y$ is not an initial segment of h. Then $P^{\pi^{h_0\mapsto m}}(h) = P^\pi(h)$.*

Proof. By induction on h. □

Lemma 51. *Suppose π, h_0, m are as in Definition 32. For any $y \in \mathcal{A}$, $P^{\pi^{h_0\mapsto m}}(h_0 y) = P^\pi(h_0)m(y)$.*

Proof. Immediate by Definition 4 and Lemma 50. □

Lemma 52. *Suppose π, h_0, m are as in Definition 32. Assume $h \in \mathcal{H}$, $y_0 \in \mathcal{A}$, and $h_0 y_0$ is an initial segment of h. Assume $\pi(y_0|h_0) \neq 0$. Then $P^{\pi^{h_0\mapsto m}}(h) = \frac{P^\pi(h)m(y_0)}{\pi(y_0|h_0)}$.*

Proof. By induction on h.

Case 1: $h = h_0 y_0$. Then

$$\begin{aligned}
P^{\pi^{h_0\mapsto m}}(h) &= P^{\pi^{h_0\mapsto m}}(h_0)\pi^{h_0\mapsto m}(y_0|h_0) && \text{(Definition 4)} \\
&= P^\pi(h_0)\pi^{h_0\mapsto m}(y_0|h_0) && \text{(Lemma 50)} \\
&= P^\pi(h_0)m(y_0) && \text{(Definition 32)} \\
&= \frac{P^\pi(h_0)\pi(y_0|h_0)m(y_0)}{\pi(y_0|h_0)} && \text{(Basic Algebra)} \\
&= \frac{P^\pi(h)m(y_0)}{\pi(y_0|h_0)}. && \text{(Definition 4)}
\end{aligned}$$

Case 2: $h = h_0 y_0 h_1 x$ for some $h_1 \in \mathcal{H}$ and $x \in \mathcal{E}$. Then

$$\begin{aligned}
P^{\pi^{h_0\mapsto m}}(h) &= P^{\pi^{h_0\mapsto m}}(h_0 y_0 h_1) && \text{(Definition 4)} \\
&= \frac{P^\pi(h_0 y_0 h_1)m(y_0)}{\pi(y_0|h_0)} && \text{(Induction)} \\
&= \frac{P^\pi(h_0 a_0 h_1 x)m(y_0)}{\pi(y_0|h_0)} && \text{(Definition 4)} \\
&= \frac{P^\pi(h)m(y_0)}{\pi(y_0|h_0)}.
\end{aligned}$$

Case 3: $h = h_0 y_0 h_1 y$ for some $h_1 \in \mathcal{H}$ and $y \in \mathcal{A}$. Then

$$\begin{aligned}
P^{\pi^{h_0 \mapsto m}}(h) &= P^{\pi^{h_0 \mapsto m}}(h_0 y_0 h_1)\pi^{h_0 \mapsto m}(y|h_0 y_0 h_1) && \text{(Definition 4)}\\
&= P^{\pi^{h_0 \mapsto m}}(h_0 y_0 h_1)\pi(y|h_0 a h_1) && \text{(Definition 32)}\\
&= \frac{P^{\pi}(h_0 y_0 h_1)\pi(y|h_0 y_0 h_1)m(y_0)}{\pi(y_0|h_0)} && \text{(Induction)}\\
&= \frac{P^{\pi}(h_0 y_0 h_1 y)m(y_0)}{\pi(y_0|h_0)} && \text{(Definition 4)}\\
&= \frac{P^{\pi}(h)m(y_0)}{\pi(y_0|h_0)}.
\end{aligned}$$

□

Proof of Proposition 37. **Subclaim:** For every $g \in \mathcal{H}$, $P^{\pi}(g) = P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g)$. We prove this by induction on g.

Case 1: $g = \varepsilon$. Then $P^{\pi}(g) = 1 = P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g)$.

Case 2: $g = fx$ for some $x \in \mathcal{E}$. Then

$$\begin{aligned}
P^{\pi}(g) &= P^{\pi}(f) && \text{(Definition 4)}\\
&= P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(f) && \text{(Induction)}\\
&= P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g). && \text{(Definition 4)}
\end{aligned}$$

Case 3: $g = fy$ for some $y \in \mathcal{A}$.

Subcase 3.1: $P^{\pi}(f) = 0$. Then

$$\begin{aligned}
P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g) &= P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(f)(\vec{w}\cdot\pi^{h\mapsto\vec{m}})(y|f) && \text{(Definition 4)}\\
&= P^{\pi}(f)(\vec{w}\cdot\pi^{h\mapsto\vec{m}})(y|f) && \text{(Induction)}\\
&= 0.
\end{aligned}$$

Similarly, $P^{\pi}(g) = 0$. So $P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g) = P^{\pi}(g)$.

Subcase 3.2: $P^{\pi}(f) \neq 0$ and $f = h$. Then:

$$\begin{aligned}
P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g) &= P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(hy)\\
&= \vec{w}\cdot P^{\pi^{h\mapsto\vec{m}}}(hy) && \text{(Theorem 14)}\\
&= w_1 P^{\pi}(h)m_1(y) + \cdots + w_n P^{\pi}(h)m_n(y) && \text{(Lemma 51)}\\
&= P^{\pi}(h)\pi(y|h) && (\vec{w}\cdot\vec{m} = \pi(\cdot|h))\\
&= P^{\pi}(hy) = P^{\pi}(g). && \text{(Definition 4)}
\end{aligned}$$

Subcase 3.3: $P^{\pi}(f) \neq 0$, $f \neq h$, and f has an initial segment hy_0 ($y_0 \in \mathcal{A}$).

Then $\pi(y_0|h) \neq 0$, lest we would have $P^{\pi}(f) = 0$. Thus:

$$\begin{aligned}
P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g) &= P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(fy)\\
&= \vec{w}\cdot P^{\pi^{h\mapsto\vec{m}}}(fy) && \text{(Theorem 14)}\\
&= w_1\frac{P^{\pi}(fy)m_1(y_0)}{\pi(y_0|h)} + \cdots + w_n\frac{P^{\pi}(fy)m_n(y_0)}{\pi(y_0|h)} && \text{(Lemma 52)}\\
&= \frac{P^{\pi}(fy)}{\pi(y_0|h)}(w_1 m_1(y_0) + \cdots + w_n m_n(y_0)) && \text{(Basic Algebra)}\\
&= \frac{P^{\pi}(fy)}{\pi(y_0|h)}\pi(y_0|h) && (\vec{w}\cdot\vec{m} = \pi(\cdot|h))\\
&= P^{\pi}(fy) = P^{\pi}(g).
\end{aligned}$$

Subcase 3.4: $P^\pi(f) \neq 0$, $f \neq h$, and f has no initial segment of the form hy_0. Then:

$$\begin{aligned}
P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(g) &= P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}(fy) \\
&= \vec{w}\cdot P^{\pi^{h\mapsto\vec{m}}}(fy) && \text{(Theorem 14)} \\
&= w_1 P^\pi(fy) + \cdots + w_n P^\pi(fy) && \text{(Lemma 50)} \\
&= P^\pi(fy) = P^\pi(g), && (w_1 + \cdots + w_n = 1)
\end{aligned}$$

as desired.

This finishes the proof of the Subclaim. By Lemma 5, the Subclaim implies that for every well-behaved μ and every $g \in \mathcal{H}$, $P^\pi_\mu(g) = P^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}_\mu(g)$. By Definition 6 (part 1), this implies that for every well-behaved μ and every $t \in \mathbb{N}$, $V^\pi_{\mu,t} = V^{\vec{w}\cdot\pi^{h\mapsto\vec{m}}}_{\mu,t}$. The proposition follows by Definition 6 (part 2). $\square$

A.8 Proof of Lemma 46

Proof. Let $h \in (\mathcal{E}\mathcal{A})^*$. Clearly $(\vec{w}\cdot\vec{\mu})(x|h) \geq 0$ for all $x \in \mathcal{E}$. It remains to show $\sum_{x\in\mathcal{E}}(\vec{w}\cdot\vec{\mu})(x|h) = 1$. If $\vec{w}\cdot P_{\vec{\mu}}(h) = 0$ then each $(\vec{w}\cdot\vec{\mu})(x|h) = 1/|\mathcal{E}|$ so the claim is immediate; assume not. Then:

$$\begin{aligned}
&\sum_{x\in\mathcal{E}}(\vec{w}\cdot\vec{\mu})(x|h) \\
&= \sum_{x\in\mathcal{E}} \frac{\vec{w}\cdot P_{\vec{\mu}}(hx)}{\vec{w}\cdot P_\mu(h)} && \text{(Definition 45)} \\
&= \sum_{x\in\mathcal{E}} \frac{\sum_{i=1}^\infty w_i P_{\mu_i}(hx)}{\sum_{i=1}^\infty w_i P_{\mu_i}(h)} && \text{(Definition 44)} \\
&= \sum_{x\in\mathcal{E}} \frac{\sum_{i=1}^\infty w_i P_{\mu_i}(h)\mu_i(x|h)}{\sum_{i=1}^\infty w_i P_{\mu_i}(h)}. && \text{(Definition 4)}
\end{aligned}$$

By absolute convergence, we can rearrange the order of summation without altering the sum, so the above is

$$\frac{\sum_{i=1}^\infty w_i P_{\mu_i}(h)\sum_{x\in\mathcal{E}}\mu_i(x|h)}{\sum_{i=1}^\infty w_i P_{\mu_i}},$$

and each $\sum_{x\in\mathcal{E}}\mu(x|h) = 1$ since each $\mu \in \Delta\mathcal{E}$, so the whole fraction reduces to 1. $\square$

A.9 Proof of Lemma 48 (1–3)

Proof. (1) By induction on h.

Case 1: $h = \varepsilon$. Then

$$\begin{aligned}
\vec{w}\cdot P_{\vec{\mu}}(h) &= \sum_{i=1}^\infty w_i P_{\mu_i}(h) && \text{(Definition 44)} \\
&= \sum_{i=1}^\infty w_i && (P_{\mu_i}(\varepsilon) = 1) \\
&= 1 && \text{(Definition of } \mathscr{W}) \\
&= P_{\vec{w}\cdot\vec{\mu}}(h). && \text{(Definition 4)}
\end{aligned}$$

Case 2: $h = gx$ for some $x \in \mathcal{E}$.

Subcase 2.1: $\vec{w}\cdot P_{\vec{\mu}}(g) = 0$. This means $\sum_{i=1}^\infty w_i P_{\mu_i}(g) = 0$. Since each $w_i > 0$, this implies each $P_{\mu_i} = 0$. From this it easily follows that $P_{\vec{w}\cdot\vec{\mu}}(gx) = \vec{w}\cdot P_{\vec{\mu}}(gx) = 0$.

Subcase 2.2: $\vec{w} \cdot P_{\vec{\mu}}(g) \neq 0$. Then

$$\begin{aligned}
P_{\vec{w}\cdot\vec{\mu}}(h) &= P_{\vec{w}\cdot\vec{\mu}}(g)(\vec{w}\cdot\vec{\mu})(x|g) && \text{(Definition 4)}\\
&= \vec{w}\cdot P_{\vec{\mu}}(g)(\vec{w}\cdot\vec{\mu})(x|g) && \text{(Induction)}\\
&= \vec{w}\cdot P_{\vec{\mu}}(g)\frac{\vec{w}\cdot P_{\vec{\mu}}(gx)}{\vec{w}\cdot P_{\mu}(g)} && \text{(Definition 45)}\\
&= \vec{w}\cdot P_{\vec{\mu}}(gx) = \vec{w}\cdot P_{\vec{\mu}}(h).
\end{aligned}$$

Case 3: $h = gy$ for some $y \in \mathcal{A}$. Then

$$\begin{aligned}
P_{\vec{w}\cdot\vec{\mu}}(h) &= P_{\vec{w}\cdot\vec{\mu}}(g) && \text{(Definition 4)}\\
&= \vec{w}\cdot P_{\vec{\mu}}(g) && \text{(Induction)}\\
&= \sum_{i=1}^{\infty} w_i P_{\mu_i}(g) && \text{(Definition 44)}\\
&= \sum_{i=1}^{\infty} w_i P_{\mu_i}(gy) && \text{(Definition 4)}\\
&= \vec{w}\cdot P_{\vec{\mu}}(gy) = \vec{w}\cdot P_{\vec{\mu}}(h). && \text{(Definition 44)}
\end{aligned}$$

(2) Compute:

$$\begin{aligned}
P^{\pi}_{\vec{w}\cdot\vec{\mu}}(h) &= P^{\pi}(h)P_{\vec{w}\cdot\vec{\mu}}(h) && \text{(Lemma 5)}\\
&= P^{\pi}(h)\vec{w}\cdot P_{\vec{\mu}}(h) && \text{(By (1))}\\
&= P^{\pi}(h)\sum_{i=1}^{\infty} w_i P_{\mu_i}(h) && \text{(Definition 44)}\\
&= \sum_{i=1}^{\infty} w_i P^{\pi}(h) P_{\mu_i}(h) && \text{(Algebra)}\\
&= \sum_{i=1}^{\infty} w_i P^{\pi}_{\mu_i}(h) && \text{(Lemma 5)}\\
&= \vec{w}\cdot P^{\pi}_{\vec{\mu}}(h). && \text{(Definition 44)}
\end{aligned}$$

(3) Let X_t, R be as in Definition 6 and compute:

$$\begin{aligned}
V^{\pi}_{\vec{w}\cdot\vec{\mu},t} &= \sum_{h\in X_t} R(h)P^{\pi}_{\vec{w}\cdot\vec{\mu}}(h) && \text{(Definition 6)}\\
&= \sum_{h\in X_t} R(h)\vec{w}\cdot P^{\pi}_{\vec{\mu}}(h) && \text{(By (2))}\\
&= \sum_{h\in X_t} R(h)\sum_{i=1}^{\infty} w_i P^{\pi}_{\mu_i}(h). && \text{(Definition 44)}
\end{aligned}$$

Since X_t is finite, this sum is absolutely convergent, so we can rearrange terms, and the sum is equal to

$$\begin{aligned}
\sum_{i=1}^{\infty} w_i \sum_{h\in X_t} R(h)P^{\pi}_{\mu_i}(h) &= \sum_{i=1}^{\infty} w_i V^{\pi}_{\mu_i,t} && \text{(Definition 6)}\\
&= \vec{w}\cdot V^{\pi}_{\vec{\mu},t}. && \text{(Definition 44)}
\end{aligned}$$

□