Theory of Collective Intelligence

David H. Wolpert
NASA Ames Research Center, Moffett Field, CA 95033
http://ic.arc.nasa.gov/~dhw

June 21, 2003

Abstract

In this chapter an analysis of the behavior of an arbitrary (perhaps massive) collective of computational processes in terms of an associated "world" utility function is presented We concentrate on the situation where each process in the collective can be viewed as though it were striving to maximize its own private utility function. For such situations the central design issue is how to initialize/update the collective's structure, and in particular the private utility functions, so as to induce the overall collective to behave in a way that has large values of the world utility. Traditional "team game" approaches to this problem simply set each private utility function equal to the world utility function. The "Collective Intelligence" (COIN) framework is a semi-formal set of heuristics that recently have been used to construct private utility functions that in many experiments have resulted in world utility values up to orders of magnitude superior to that ensuing from use of the team game utility. In this paper we introduce a formal mathematics for analyzing and designing collectives. We also use this mathematics to suggest new private utilities that should outperform the COIN heuristics in certain kinds of domains. In accompanying work we use that mathematics to explain previous experimental results concerning the superiority of COIN heuristics. In that accompanying work we also use the mathematics to make numerical predictions, some of which we then test. In this way these two papers establish the study of collectives as a proper science, involving theory, explanation of old experiments, prediction concerning new experiments, and engineering insights.

Introduction

This paper concerns distributed systems some of whose components can be viewed as though they were agents, adaptively "trying" to induce large values of their associated private utility functions. When combined with a world utility function that rates the possible behaviors of that system, the system is known as a **collective** [17, 20, 23, 25].

Given a collective, there is an associated inverse design problem, of how to configure/modify the system so that in their pursuit of their private utilities the agents also maximizes the world utility. Solving this problem may involve determining/modifying the number of agents, how they interact with each other, and what degrees of freedom of the overall system each of them controls (i.e., the very definition of the agents). When the agents are machine learning algorithms overtly trying to maximize their private utilities, the inverse problem may also involve determining/modifying the algorithms that those agents use, as well as precisely what private utilities they are each trying to maximize.

This paper presents a mathematical framework for the investigation of collectives, and in particular the investigation of this design problem. A crucial feature of this framework is that it involves no modeling of the underlying system nor of the algorithms controlling the agents. For example, only the behavior of an agent (or more precisely, certain broad aspects of it) is formally related to what private utility that agent is "trying" to maximize; nothing of what goes on "under the hood" is assumed. This behaviorist approach is crucial since in the real world collectives are often so complicated that no tractable model can bear more than a cursory similarity with the system it is supposed to represent. More generally, this approach is crucial to have the framework be broad enough to encompass, for example, the collectives of spin glasses and of human economies.

In the next section we introduce generalized coordinates. These allow us to avoid any restrictions on the kinds of variables comprising the system—they can be uncountable, countable, or combinations thereof, with or without an underlying topology/metric, and except where explicitly indicated otherwise, all the results of the framework still apply. The underlying variables can either include time or not, and if they do, the associated underlying dynamics is arbitrary. The variables also can either be broken up explicitly into separate agents or not, and if they are, there can be arbitrary restrictions on which of the conceivable joint moves of the agents are physically allowed. In addition, how the variables are broken up into agents, and even the number of agents is arbitrary, and can be modified dynamically (if time is included in the underlying variables). Moreover, if time is included as an underlying variable, then some of the agents can have their decision "simultaneously" fix the state of one or more variables of the system *at distinct moments in time.* (This is reminiscent of what is decided in settling on a contract in cooperative game theory.) Again, all of this can be varied in an arbitrary fashion.

Using these generalized coordinates, a central equation can be derived that determines how well any of these kinds of systems perform. It does so by breaking performance down into three terms. These terms loosely reflect the concerns of the fields of high-dimensional search, economics, and machine learning; the central equation is the bridge that couples those fields.

The following section uses this mathematical framework to introduce a (model-independent) formalization of the assumption that a particular component of the system is a "utility-maximizing... agent". That formalization is then used to derive the Aristocrat and Wonderful Life private utility functions, two utility functions previously intuited that have been found to result in far better world utility than conventional techniques. [17]. This derivation also uncovers (relatively rare) conditions under which those utilities should not perform very well. That section ends by deriving many new results, including the Collapsed private Utility, and ways to modify other agents to help a particular agent, along with specification of the scenarios in which such techniques should result in good world utility.

An accompanying paper [22] presents this mathematical framework in a more pedagogical manner, including many examples, commentary and some discussion of related fields (e.g., mechanism design in game theory). That paper also discusses recent experiments involving a set of previous semi-formal heuristics (including the Aristocrat and Wonderful Life private utilities) that have been found to be very useful for the design of collectives. It uses the mathematical framework to explain the efficacy of those techniques. It then goes on to make numerical predictions based on that framework, and then presents some experimental tests of those predictions. It ends by making other (testable) predictions, and presents a sample of future research topics and open issues.

This paper instead exhaustively presents all of the currently elaborated mathematics of the framework, including the details omitted in [22]. In particular, this paper contains theorems not presented there, extensions of the theorems that are presented there, the proofs of all theorems, detailed application of the framework to multi-step games, and the important example of applying the framework to gradient ascent over categorical variables. (For pedagogical reasons, the latter two occur as appendices.) Combined, these two papers present a mathematical theory along with associated predictions/experiments and engineering recommendations. In this, they lay the foundation for a full-fledged science of collectives.

1 The Central Equation

(i) Generalized coordinates and intelligence

We are interested in addressing optimization problems by decomposing them into many subproblems, each of which are solved separately. We will not try to choose such subproblems so that they are independent of one another, or find a way to coordinate their solutions. Rather we will choose the subproblems so that each of them separately is relatively easy to solve, *given the context of a particular current solution to the other subproblems*, and then have them be solved in parallel.

To formalize this, let ζ be an arbitrary space with elements z called **worldpoints.** Let $C \subseteq \zeta$ be the set of elements of ζ that are actually allowed, for example in that they are consistent with the laws of physics.[1] Define a **generalized coordinate variable** as a function from C to associated **coordinate**

[1] Whenever expressing a particular system as a collective, it is a good rule to write out the functional dependencies presumed to specify $C(.)$ as explicitly as one can, to check that what one has identified as the space ζ does indeed contain all the important variables.

values. (When the context makes the precise meaning clear, we will sometimes use the term "coordinate" to refer to a generalized coordinate variable, and sometimes to a value of that variable.) We will sometimes view a coordinate variable ρ as an exhaustive partition of C into non-empty subsets, with $\rho(z)$ being the element of the partition that contains z. Accordingly we will sometimes write a coordinate value $r = \rho(z)$ as "$r \in \rho$" and a worldpoint z' sharing that value as "$z' \in r$".[2] Intuitively, each "sub-problem" of our overall optimization problem will be formalized in terms of such a partition ρ, as finding the optimal z within the $r \in \rho$ specified by the current solutions to the other subproblems.

Often we implicitly assume that the set of values that any coordinate variable we are discussing can take on forms a measurable set, as does the set of worldpoints having any such value. (All integrals are implicitly with respect to such measures.)

As an example, C might consist of the possible joint actions of a set of computational agents engaged in a non-cooperative game [7, 2, 10, 3, 5]. $\rho(z \in C)$ could then be the actions of all agents *except* some particular agent identified with ρ. In this case, by fixing all other degrees of freedom, the value of the coordinate ρ implicitly specifies the degrees of freedom that are still "available to be set" by the agent identified with ρ.

A frequently occurring type of coordinate variable is one whose values are contained in the real numbers. A particularly important example is a **world utility** function $G : C \rightarrow \mathfrak{R}$ that ranks the various possible worldpoints of the system. We are always provided a G; the goal in the problem of designing collectives is to maximize G.

Our mathematics does not concern G alone, but rather its relationship with some **coordinate utilities** $g_\rho : C \rightarrow \mathfrak{R}$.[3] Each coordinate utility ranks the possible values of those degrees of freedom still allowed once the worldpoint has been restricted to a set of worldpoints $r \in \rho$. Given a set of coordinate variables, $\{\rho\}$, we are interested in inducing a z that each g_ρ ranks highly (relative to the other worldpoints in the associated set $r = \rho(z)$), and in the relation between those rankings of z and G's ranking of z. To analyze these issues we need to standardize utility functions so that the numeric value they assign to z only reflects their relative ranking of z (potentially just in comparison to the other worldpoints sharing some associated coordinate value).[4]

Generically, we indicate such a standardization by N, and for any utility function U, coordinate ρ, and $z \in C$, we write the associated value of such a standardization of the utility U as $N_{\rho,U}(z)$. Define "sgn[x]" to equal $+1$, 0, or -1 in the usual way. Then we only need to require of a standardization N that $N_{\rho,U}(z)$ be a $[0, 1]$-valued, ρ-parameterized functional of the pair $(U, U(z))$, one that meets the following two conditions as we vary U and/or z:

[2] In general, we try to use lower-case greek letters for coordinates, and the associated lower-case roman letter for the value of that coordinate.

[3] In previous work, roughly analogous utilities were called "personal utilities" [17].

[4] It turns out that there never arises a reason to consider the relation between such a standardization and the axioms conventionally used to derive utility theory [10], and in particular those axioms concerning behavior of expectation values of utility.

i) $\forall\, z \in C$, if for a pair of utilities V and W, $\text{sgn}[W(z')-W(z)] = \text{sgn}[V(z')-V(z)]\ \forall\, z' \in \rho(z)$, then $N_{\rho,W}(z) = N_{\rho,V}(z)$.

ii) With U and $r \in \rho$ fixed, $\forall\, z, z' \in r$, $\text{sgn}[N_{\rho,U}(z) - N_{\rho,U}(z')] = \text{sgn}[U(z) - U(z')]$.

We call the value of $N_{\rho,U}$ at z the "**intelligence** of z (given ρ) with respect to U for coordinate ρ".[5] ,[6] If ρ consists of a single set (all of C), we simply write $N_U(z)$. An example of an intelligence operator based on percentiles is provided in App. A. Unless explicitly stated otherwise, whenever calculating intelligence values in any examples, we will use this choice of the intelligence operator.

Often there will be uncertainly in the worldpoint z, in particular on the part of the system designer (e.g., when worldpoints are worldlines of a physical system, such uncertainty arises if the designer is not able to calculate exactly how the system evolves). Such uncertainty is captured by a distribution $P(z)$ that equals 0 off of C.[7] Accordingly, coordinates ρ are not only partitions, but are also random variables, taken values $r \in \rho$.

All aspects of the designer's ability to manipulate the system are encapsulated in the selection of an element s from some **design coordinate** σ. In particular, since the (sub)problem of finding a $z \in r$ with maximal ρ-intelligence will vary as r varies, it cannot be addressed with conventional algorithms for maximizing a static function. Instead, its solution requires techniques — like those in reinforcement learning — tailored for dynamically varying and/or uncertain functions. Accordingly, we will often consider the case where (among other things) s specifies which of a set of allowed **private utility** functions to associate with some coordinate $\rho, \underline{g}_{\rho,s} : z \rightarrow \mathfrak{R}$. Such a function is one that we view intuitively as the "payoff function" for a self-interested computational

[5] Note that for fixed U, the function $N_{\rho,U}(.)$ from $C \rightarrow \mathfrak{R}^+$ can be viewed as a utility function, and therefore as a coordinate. In particular, $N_{\rho,N_{\rho,U}} = N_{\rho,U}$. This follows from condition (i) in the definition of intelligence with $V = U$, $W = N_{\rho,V}$, and the equality of sgn's following from condition (ii) in the definition of intelligence.

[6] Although this paper concentrates on $\mathfrak{R}$-valued utility functions, much of its analysis can be extended to functions having different ranges. Examples include vector-valued functions having range $\mathfrak{R}^n$ — appropriate for analyzing intelligence with respect to several distinct U at once — and functions whose range is a set of non-overlapping contiguous sub-intervals of $\mathfrak{R}$. In particular, given some such range Q, and any associated antisymmetric preference function $F : Q \times Q \rightarrow \{-1, 0, 1\}$, we can replace the sgn function with F throughout (i) and (ii) when we specify our intelligence operator. Much of the sequel (e.g., Thm. 1) still holds under this modification. If in addition Q is a field over the reals, we can also form the average value of such an intelligence, and some of the theorems presented below concerning expected intelligence values will go through.

[7] If there is uncertainty in C itself we express that with a distribution $P(C)$, to go with the distributions $P(z \mid C)$. In particular, if probabilities reflect the system designer's uncertainty about C, then $P(z)$ may be non-zero even for points z off of the actual C. Fixing C exactly is analogous to fixing the energy exactly in statistical physics (the microcanonical ensemble), with allowing C to vary being analogous to uncertainty in the energy (the canonical ensemble). Unless explicitly stated otherwise, in this paper we will consider C to be fixed. In a similar fashion, if probabilities reflect uncertainty in how a coordinate κ partitions C, then it could be that $P(z \mid k)$ is non-zero even for points z where $\kappa(z) \neq k$. (For simplicity, we will usually assume this is not the case.)

agent, embodied in C, that uses a "learning algorithm", to "control" position within any particular element of ρ.[8] *A priori*, a coordinate need not have an associated private utility; in particular, non-learning agents need not. Informally, when we have a "learning agent" associated with coordinate ρ we refer to ρ as either the **agent coordinate** or the agent's **context coordinate**, with the value of that coordinate being the agent's **context**. (These definitions are made more formal below.)

Properly interpreted, the rules of set theory hold when coordinate variables play the role of sets. Under this interpretation any coordinate variable κ arising in a set-theoretic expression should be read as "every (subset of ζ that constitutes an) element of κ". For example, $\kappa \subset \lambda$ means "every element of κ is a proper subset of every element of λ", so that the value k fixes l. See App. B.

As a notational matter, we adopt the usual convention that probability of a coordinate value is shorthand that the associated random variable takes on that value, e.g., $P(a)$ means $P(\alpha = a)$. As usual though, *this convention is not propagated to expectation values*: $E(U(a,\beta) \mid c) \equiv \int \mathrm{d}b U(a,b) P(b \mid c)$. Delta functions are either Kronecker or Dirac as appropriate (although always written as arguments rather than as subscripts). Similarly, integrals are assumed to have a point-mass measure (i.e., reduce to a sum) as appropriate. For any function $\phi : C \rightarrow \Re$ and coordinate κ, with $y \in [0,1]$, we write $\mathrm{CDF}_\phi(y \mid k)$ to mean the cumulative distribution function $P(\phi \leq y \mid k) \equiv \int_{-\infty}^{y} \mathrm{d}t \int \mathrm{d}z\, P(z \mid k)\, \delta(\phi(z) - t)$, and just write $\mathrm{CDF}(\phi \mid k)$ to refer to the entire function over y. In addition, "supp" is shorthand for the support operator, and "$\mathfrak{B}$" indicates the Booleans. $O(A)$ means the cardinality of the set A. For any two functions f_1 and f_2 with the same domain $x \in X$, "$f_1 < f_2$" means that $\forall x \;\; f_1(x) \leq f_2(x)$, and $\exists x$ such that $f_1(x) < f_2(x)$. All proofs that are not in the text are provided in App. C.

(ii) The Central Equation

Our analysis revolves around the following **central equation** for $P(U \mid s)$, which follows from applying Bayes' theorem twice in succession:

$$P(U \mid s) = \int \mathrm{d}\vec{N}_U\, P(U \mid \vec{N}_U, s) \int \mathrm{d}\vec{N}_g\, P(\vec{N}_U \mid \vec{N}_g, s) P(\vec{N}_g \mid s) \qquad (1)$$

where usually we are interested in having $U = G$. "g" is the vector of the values of a set of coordinate utilities, and "$\vec{N}_g$" is an associated vector of intelligences with respect to those coordinate utilities. Here we concentrate on the case where each of those intelligences is for the associated coordinate, i.e., for set of coordinates $\{\rho\}$ it is the ρ-indexed vector with components $\{\vec{N}_{\rho,g_\rho}(z)\}$. "$\vec{N}_U$" is also a coordinate-variable-indexed vector of intelligence values, only for utility U. We will concentrate on the case where $\vec{N}_U$ is indexed with the same coordinates as $\vec{N}_g$. In this situation $\vec{N}_U$ has components $\vec{N}_{\rho,U}(z)$ and is identical to $\vec{N}_g$ except

[8] Note that, formally speaking, the learning algorithm itself is embodied in C. Hence the quotation marks around the term 'control'.

in its choice of utility functions.[9]

If we can choose s so that term 3 in the integrand in Eq. 1 is peaked around vectors $\vec{N}_g$ all of whose components are close to 1, then we have likely induced large intelligences. If in addition to such a good term 3 we can have term 2 be peaked about $\vec{N}_U$ equal to $\vec{N}_g$, then $\vec{N}_U$ will also be large. If in addition term 1 in the integrand is peaked about high U when $\vec{N}_U$ is large, then our choice of s will likely result in high U, as desired.

In the next subsection we analyze what coordinate utilities give the desired form of term 2 in the central equation, for our choice of $\vec{N}_G$ and $\vec{N}_g$. We then present examples illustrating such systems and more generally illustrating generalized coordinates. We end this section with a brief discussion of term 1. Then in the next section we analyze what coordinate utilities give the desired form of term 3 in the central equation. It is only here that the use of agents to control some coordinate values becomes crucial. We end that section by combining these analyses to derive coordinate utilities that have the desired forms for both term 2 and term 3.

This formalism applies to many more scenarios than those that involve dynamical systems with values z specifying behavior across time. It also applies even in scenarios that are not conventionally viewed as instances of game theory. Nonetheless, as an example of the formalism, App. D is a detailed exposition of multistep games in terms of this formalism.

(iii) Term 2—Factoredness

We say that U_1 and U_2 are (mutually) **factored** at a point z for coordinate ρ if $N_{\rho,U_1}(z') = N_{\rho,U_2}(z')\ \forall\ z' \in \rho(z)$.[10] Note that factoredness is transitive. If we do not specify U_2, it is taken to be G, and we sometimes say that U "is factored", or "is factored with respect to G", when U and G are mutually factored. If $\forall\ \rho$ in a set of coordinates that we are using to analyze a system, the utility g_ρ is factored with respect to G for coordinate ρ at a point z, we simply say that the system is factored at z, or that the $\{g_\rho\}$ are factored with respect to G there.

There is a very tight relation between factoredness and game theory. For example, consider the case where we have Pareto superiority of a point z' over some other point z with respect to the coordinate utility intelligences [7, 2, 10, 3, 5]. Say that in addition those associated utilities form a factored system with respect to the world utility G. These together imply the Pareto superiority of z' over z with respect to world utility. The converse also holds. However these properties relating factoredness, coordinate and world utilities only hold for Pareto superiority for intelligences (rather than for raw coordinate utility values), in

[9] Since the distributions in Eq. 1 are conditioned on s, when we have a percentile-style intelligence, a natural choice for the associated measure $d\mu(z)$ is given by the values $r = \rho(z)$ and s, as $P(z \mid r)P(r \mid s)$ (see App. A). In other words, given that we are within a particular r, the measure extends across that entire context—including points inconsistent with s—according to the distribution $P(z \mid r)$.

[10] In previous work we defined factoredness only to mean that $\text{sgn}[U_1(z') - U_1(z)] = \text{sgn}[U_2(z') - U_2(z)]\ \forall\ z' \in \rho(z)$. This is a necessary (but not sufficient) condition that $N_{\rho,U_1}(z') = N_{\rho,U_2}(z')\ \forall\ z' \in \rho(z)$; see Thm. 1 below and the definition of intelligence.

general. In addition, by taking $U_2 = G$, the following theorem provides the basis for relating game-theoretic concepts like Nash equilibria and non-rational behavior with world utility in factored systems:

Theorem 1 *U_1 and U_2 are mutually factored at $z \in C$ for coordinate ρ iff*

$$\mathrm{sgn}[U_1(z') - U_1(z'')] = \mathrm{sgn}[U_2(z') - U_2(z'')] \qquad \forall\, z', z'' \in \rho(z).$$

Note that this holds regardless of the precise choice of N, so long as it meets the formal definition of an intelligence operator.

By Thm. 1, for a system whose coordinate utilities are factored with respect to G, the set of Nash equilibria of those coordinate utilities equals the set of points that are maxima of the world utility along each of the coordinates individually (which of course does not mean that they are maxima along off-axis directions).[11] In addition to this desirable equilibrium structure, factoredness ensures the appropriate off-equilibrium structure; so long as for each coordinate the associated intelligence is high (with respect to that coordinate's utility), the system will be close to a local maximum of world utility. This is because, for each coordinate ρ, given a (fixed) associated coordinate value r, any change in $z \in r$ that decreases ρ's coordinate utility—which is almost all changes if ρ's intelligence is high—will assuredly decrease world utility. Note though that having g_ρ factored with respect to G does not preclude deleterious side-effects on the other coordinate utilities of such a g_ρ-improving change within r. All such factoredness tells us is whether world utility gets improved by such changes (see the end of App. D).[12]

[11] An immediate game-theoretic corollary is that any game whose utilities can be expressed as coordinate utilities of a system that is factored with respect to a world utility having critical points has at least one pure strategy Nash equilibrium. However consider an arbitrary vector $\vec{\varepsilon}$ all of whose components lie in $[0,1]$. Then it is not the case that every factored system has a pure strategy joint profile with each player's intelligence given by the associated component of $\vec{\varepsilon}$. This is even true if every component of $\vec{\varepsilon}$ is either a 0 or a 1. As a simple example, choose $g_1 = g_2 = G$, and have $\vec{\varepsilon} = (0,1)$. Have $G = z_1$ for $z_2 > 1/2$, and equal $1 - z_1$ otherwise, where both z_1 and $z_2 \in [0,1]$. Then if $z_2 > 1/2$, $z_1 = 1$, since $N_1 = 1$. However if $z_1 = 1$, then $z_2 \in [0, 1/2]$ since $N_2 = 0$. If $z_2 \leq 1/2$ though, $z_1 = 0$, which means that $z_2 \in (1/2, 1]$. **QED.**

[12] Factoredness is simply a bit; a system is factored or it isn't. As such it cannot quantify situations in which term 2 has a good form although it is not exactly a delta function. Nor can it characterize "super-factored" situations in which that conditional distribution is *better* than a delta function, being biased towards N_G values that exceed the N_g values. One way to address this deficiency is to define a "degree of factoredness". One example of such a measure is $1 - \int \mathrm{d}z\, P(z \mid s)[\vec{N}_G - \vec{N}_g]^2 \in [0,1]$. Another is $\int \mathrm{d}z\, P(z \mid s)[\vec{N}_G - \vec{N}_g]$, which extends from "partially factored" systems (negative values), to perfectly factored systems (value 0), to super-factored systems (value greater than 0). Other definitions arise from consideration of Thm. 1. For example, one might quantify factoredness for coordinate ρ as the probability that a random move within a context changes G and g_ρ the same way:

$$\int \mathrm{d}z\, \mathrm{d}z'\, P(z \mid s) P(z' \mid s) \delta(z' \in \rho(z)) \Theta([G(z) - G(z')][g_\rho(z) - g_\rho(z')]).$$

Especially when one has a percentile-type intelligence, all these possibilities suggest yet other variants in which the measure $\mathrm{d}\mu(z)$ replaces the distribution(s) $P(z \mid s)$. Similarly, one can define "local" (degree of factoredness) about some point z'' by introducing into the integrands of all these variants Heaviside functions restricting the worldpoint to be near z''.

The following theorem gives the entire equivalence class of utilities that are mutually factored at a point:

Theorem 2 *U_1 and U_2 are mutually factored at z for coordinate ρ iff $\forall\ z' \in r \equiv \rho(z)$, we can write*

$$U_1(z') = \Phi_r(U_2(z'))$$

for some r-indexed function Φ_r that is a strictly increasing function of its argument across the set of all values $U_2(z' \in r)$. (The form of U_1 for other arguments is arbitrary.)

Using some notational overloading of the "Φ" function, by Thm. 2 we can ensure that the system is factored by having each $g_\rho(z) = \Phi_\rho(G(z), \rho(z))\ \forall\ z \in \zeta$ for some functions Φ_ρ whose first partial derivative is strictly increasing everywhere. Note that this factoredness holds regardless of C or $P(z \mid s)$. The canonical example of such a case is a **team game** (also known as an 'exact potential game' [6, 12, 4]) where $g_\rho = G$ for all ρ. Alternatively, by only requiring that $\forall\ z \in C$ does g_ρ take on such a form, we can access a broader class of factored utilities, a class that *does* depend on aspects of C.

As an example, define a **difference** utility for coordinate ρ with respect to utility D_1 as a utility taking the form $D_\rho(z) = \beta(z)[D_1(z) - D_2(z)]$ for some function D_2 and positive function $\beta(.)$, where both $\beta(.)$ and $D_2(.)$ have the same value for any pair of points z and $z' \in C$ for which $\rho(z) = \rho(z')$. (We will sometimes refer to D_1 as the **lead utility** of such a difference utility, with D_2 being the **secondary utility**.) Since both $\beta(z)$ and $D_2(z)$ can be written purely as a function of $\rho(z)$, by Thm. 2, a difference utility is factored with respect to D_1. As explicated in the next subsection, for such a utility with $D_1 = G$, term 3 in the central equation can be vastly superior to that of a team game, especially in large systems. In addition, as a practical matter, often D_ρ can be evaluated much more easily than can D_1.

(iv) Term 1 and alternate forms of the central equation

Assuming term 3 results in a large value of $\vec{N}_g$, having factoredness then ensures that we have a large value of $\vec{N}_G$ as well. In this situation term 1 will determine how good G is. Intuitively, term 1 reflects how likely the system is to get caught near local maxima of G. If any maximum of G the system finds is likely to be the global maximum, then term 1 has a good form. (For factored systems, in such scenarios it is likely that a system near a Nash equilibrium it is near the highest possible G.)

So for factored systems, for our choice of $\vec{N}_G$ and $\vec{N}_g$, term 1 can be viewed as a formal encapsulation of the issue underpinning the much-studied exploration/exploitation trade-off of conventional search algorithms. That trade-off can manifest itself both within the learning algorithms of the individual agents as well as in a centralized process determining whether those agents are allowed to make proposed changes in their state ([26]). In this paper we will not consider such issues, but will instead concentrate on terms 2 and 3.

As mentioned, term 2 in the central equation is closely related to issues considered in economics and game theory (cf. Thm. 1 and note the relation between factoredness and the concept of incentive compatibility in mechanism design [7, 2, 14, 2, 10, 16, 8, 27, 13, 15]. On the other hand, as expounded below, term 3 is closely related to signal-noise issues often considered in machine learning (but essentially never considered in economics). Finally, as just mentioned, term 1 is related to issues considered by the search community. So the central equation can be viewed as a way of integrating the fields of economics, machine learning, and search.

Finally, an important alternative to the choice of $\vec{N}_U$ investigated in this paper is where it is the scalar $N_{\emptyset,U}$. In this situation, $\vec{N}_U$ is a monotonic transformation of U *over all of* C, rather than just within various partition elements of C. For this choice term 1 in the central equation becomes moot, and that equation effectively reduces to $P(U \mid s) = \int \mathrm{d}\vec{N}_g P(U \mid \vec{N}_g, s) P(\vec{N}_g \mid s)$. The analysis presented below of the $P(\vec{N}_g \mid s)$ term in the central equation is unchanged by this change. However the analysis of the $P(\vec{N}_U \mid \vec{N}_g, s)$ term is now replaced by analysis of $P(U \mid \vec{N}_g, s)$. For reasons of space, we do not investigate this alternative choice of $\vec{N}_U$ in this paper.

2 The Three Premises

(i) Coordinate complements, moves, and worldviews

Since intelligence is bounded above by 1, we can roughly encapsulate the quality of term three in the central equation as the associated expected intelligence. Accordingly, our analysis of term 3 will be expressed in terms of expected intelligences.

We will consider only one coordinate at a time together with the associated expected coordinate intelligence. This simplifies the analysis to only concern one of the components of $\vec{\varepsilon}_g$ together with the dependence of that component on associated variations in s, our choice of the element of the design coordinate. For now we further restrict attention to agent coordinate utilities, reserve "ρ"to refer only to such an agent coordinate with some associated learning algorithm, and take $g_\rho = \underline{g}_{\rho,s}$.[13] The context will always make clear whether ρ specifies a coordinate (as when it subscripts a private utility), refers to the values the coordinate can assume (as in $r \in \rho$), indicates the associated random variable (as in expressions like $P(U(x, \rho)) = \int \mathrm{d}r P(r) U(x, r)$), etc.

As a notational matter, define two partitions of some $T \subseteq \zeta$, π_1 and π_2, to be **complements** over $T \subseteq \zeta$ if $z \in T \rightarrow (\pi_1(z), \pi_2(z))$ is invertible, so that,

[13] Note that changing ρ's coordinate utility while leaving s unchanged has no effect on the probability of a particular G value; g_ρ is just an expansion variable in the central equation. Conversely, leaving ρ's coordinate utility the same while making a change to its private utility (and therefore to s, and therefore in general to the associated distribution over ζ, $P(z \mid s)$) changes the probability distribution across G values. Setting those two utilities equal is what allows the expansion of the central equation to be exploited to help determine s.

intuitively speaking, π_1 and π_2 jointly form a "coordinate system" for T.[14,15] When discussing generalized coordinates, this nomenclature is used with T implicitly taken to be C. (π_1 and π_2 are coordinate variables in the formal sense if $T = C$.) We adopt the convention that for any coordinate ρ, $\hat{}\rho$, having labels/values written $\hat{}r$, is shorthand for some coordinate that is complementary to ρ (the precise such coordinate will not matter) and that $\hat{}\hat{}\rho = \rho$. We do not take the "$\hat{}$" operator to refer to values of a coordinate, only to coordinates as a whole. So for example, there is no *a priori* relationship implied between a particular element of $\hat{}\rho$ that we write as "$\hat{}r$", and some particular element of ρ that we write as "r".

We always have $E(N_{\rho,U} \mid s) = \int drdndx P(r \mid s) P(n \mid r, s) P(x \mid n) N_{\rho,U}(x, r)$. Accordingly, if we knew $P(r \mid s)$, and also knew one of $P(n \mid r, s)$ and $P(x \mid n)$ but did not know the other, then we could in principle solve for that other distribution so as to optimize expected intelligence.[16] Unfortunately, we usually do not know two of those three distributions, and so must take a more indirect approach.

The analysis presented here for agent coordinates revolves around the issue of how sensitive g_ρ is to changes within an element of ρ as opposed to changes between those elements of ρ. To conduct this analysis we will need to introduce two coordinates in addition to σ and ρ: ξ and ν.[17] Given some $\hat{}\rho$, rather than the precise element $\hat{}r \in \hat{}\rho$, in general the agent associated with ρ can only control which of several sets of possible elements $\hat{}r$ the system is in. This is formalized with the coordinate $\xi \supseteq \hat{}\rho$. We refer to ξ as the **move variable** of the agent, and we refer to an $x \in \xi$, and/or the set of z that that x specifies, as the **move value** of the agent. For convenience we assume that for all such contexts r and moves x there exists at least one $z \in C$ such that $\rho(z) = r$ and $\xi(z) = x$. In general, what we identify as the ξ of a particular ρ need not be unique. Intuitively, such a partition ξ delineates a set of $r \rightarrow z$ maps, each such map giving a way that the agent associated with ρ is allowed to vary its behavior to reflect what context r it's in. An agent's move is a selection among such a set of allowed variations. An important example of move variables involving dynamic processes in presented in App. D.

We assume that $\xi(z)$ and $\rho(z)$ jointly set the value of $G(z)$ and of any $\underline{g}_{\rho,s}$ we will consider.[18] Accordingly, we write $\underline{\gamma}_\rho$ when we mean the coordinate whose partition elements are identical to σ's but whose values are instead the private

[14] This characterization as a coordinate system is particularly apt if π_1 and π_2 are **minimal** complements, by which is meant that there is neither a coarser partition $\pi' \supseteq \pi_1$ such that π' and π_2 are complements, nor a coarser partition $\pi'' \supseteq \pi_2$ such that π'' and π_1 are complements.

[15] Note that it is not assumed that $T \rightarrow (p_1, p_2)$ taking points z to partition element pairs is surjective.

[16] Formally, to implement this would require making an associated change to s, a change which in the case of solving for $P(x \mid n)$ would have to be reflected in the value of n.

[17] Properly speaking, ξ and ν should be indexed by ρ, as should the coordinates $\sigma_{\underline{g}}$ and $\sigma_{\hat{}\underline{g}}$ introduced below; for reasons of clarity, here all such indices are implicit.

[18] Phrased differently, given the utility function, and the associated ξ and ρ, the minimal choice for ζ is $\xi \times \rho$. If the value s is not fixed by $x \times r$, i.e., if it is not the case that $\sigma \supseteq \xi \cap \rho$, then σ must also be contained in ζ, and similarly for ν.

utility functions of ρ: $\underline{\gamma}_\rho : s \in \sigma \rightarrow \underline{g}_{\rho,s}$. Similarly, we will write N_ρ when we mean the function $(x, r, s) \rightarrow N_{\rho,\underline{g}_{\rho,s}(x,r)}$.

We refer to ν as the **worldview variable** of the agent, and we refer to a $n \in \nu$, and/or the set of possible z that that ν specifies, as the **worldview value** of the agent. Intuitively, n specifies all the information—all training data, all knowledge of how the training data is formed (including potentially knowledge of its own private utility), all observations, all external commands, all externally set prior biases—that ρ's agent uses to determine its move, and nothing else. It is the contents of the (perhaps distorting) "window" through which the learning algorithm receives information from the external world.

Formally, there three properties a coordinate must possess for it to qualify as a worldview of an agent. First, if the agent does indeed use all the information in n, then the agent's preference in moves must change in response to any change in the value of n. This means that $\forall\ n_1, n_2 \in \nu$, for at least one of the $x \in \xi$, $P(x \mid n_1) \neq P(x \mid n_2)$.[19] Second, if the worldview truly reflects everything the agent uses to make its move, then any change to any variable must be able to affect the distribution over moves only insofar as it affects n. This means that with Ω defined as the set of all non-ξ coordinate we will consider in our analysis (e.g., σ, ρ for some other agent, their intersection, etc.), $P(x \mid n, W) = P(x \mid n)\ \forall\ x \in \xi$, $n \in \nu$ and $W \in \Omega$ such that $P(x, n, W) \neq 0$.[20,21,22] Finally, of all coordinates obeying these two properties, the worldview must be among those whose information maximizes the expected performance of the associated Bayes-optimal guessing,[23] i.e., $\forall\ s \in \sigma, \beta \neq \nu$,

$$\begin{aligned}\int \mathrm{d}b P(b \mid s) E\big(\ &\underline{\gamma}_\rho[\mathrm{argmax}_{\mathrm{x}'}\{E(\underline{\gamma}_\rho(x', \rho) \mid b)\}, \rho]\ \mid b\ \big)\\ &\leq \int \mathrm{d}n P(b \mid s) E\big(\ \underline{\gamma}_\rho[\mathrm{argmax}_{\mathrm{x}'}\{E(\underline{\gamma}_\rho(x', \rho) \mid n)\}, \rho]\ \mid n\ \big).\end{aligned}$$

So $P(n \mid s)$ is how the worldview varies with s, and $P(x \mid n)$ is how the agent's learning algorithm uses the resultant information. The $P(x \mid s)$ induced by these two distributions is how the move of the agent varies with s. Alternatively, $P(r \mid s)$ is the distribution over contexts caused by our choice of design coordinate value, and the distribution $P(x \mid r, s) = \int \mathrm{d}n P(x \mid n) P(n \mid r, s)$ gives all salient aspects of the agent's learning algorithm and technique for inferring information abou r; the integral over r of the product of these two distributions says how choice of s determines the distribution over moves.

[19] When worldviews are numeric-valued, we can modify this requirement to be that the distribution $P(x \mid n)$ has to be sufficiently sensitive a function of n over all of ν.

[20] Note that if *all* W are allowed, then in general the only choice for ν obeying this restriction is $\nu = \xi$.

[21] As a result of this requirement, $P(r \mid x, n, W) = P(r \mid n, W)$, $P(x, r \mid n, W) = P(x \mid n) P(r \mid n, W)$, etc.

[22] For any $P(z)$ and coordinates α and β , one can always construct a coordinate $\delta \neq \alpha$ such that $P(a \mid b, d)$ varies with d. So our assumption about ξ, ν and Ω constitutes a restriction on what coordinates we will consider in our analysis.

[23] If it were not for this requirement, ξ could double as the worldview, and often so could σ.

We will find it convenient to decompose $\sigma = \sigma_{\underline{\gamma}_\rho} \cap \sigma_{\hat{\underline{\gamma}}_\rho}$, where $\sigma_{\underline{\gamma}_\rho}$ is a coordinate whose value gives $\underline{g}_{\rho,s}$, and there is no coordinate $\omega \supset \sigma_{\underline{\gamma}_\rho}$ with this property. (Intuitively, $\sigma_{\underline{\gamma}_\rho}$'s value is a component of s that specifies $\underline{g}_{\rho,s}$ and nothing more.) Also, from now on, we will often drop the ρ index whenever its implicit presence is clear. So for example, we will often write $s_{\underline{g}}$ instead of $s_{\underline{\gamma}_\rho}$.

(ii) Ambiguity

Since we do not know $P(x \mid n)$ in general, we cannot directly say how n sets the distribution over x. Fortunately we do not need such detailed information. We only need to know the effect that certain changes to n have on particular characteristics of the associated distribution $P(x \mid n)$ (e.g., the effect certain changes to n have on the "characteristic of $P(x \mid n)$" given by an n-conditioned expected intelligence $E(N_U \mid n)$).

Now if there were any universal rule for how such characteristics affect expected intelligence, then without any assumptions we could use such a rule to deduce that some particular choices of n are superior to others. That has been proven to be impossible however [18, 21]. Accordingly, we must make some presumption about the nature of the learning algorithm, one that must be as conservative as possible if it is to apply to all reasonable algorithms.

To see what presumption we can safely make concerning such effects, first note that the worldview n encapsulates all the information the agent might try to exploit concerning the x-dependence of the likely values of the private utility. That encapsulation given by n takes the form of the distribution over the Euclidean vector of private utility values $(y^1, y^2, ...)$ given by $\int drds\ \delta(\underline{g}_{\rho,s}(x^1, r) - y^1)\delta(\underline{g}_{\rho,s}(x^2, r) - y^2)...\ P(r, s \mid n)$. The agent works by "trying" to use this encapsulation to appropriately set its move. Our presumption must concern aspects of how it does this. Furthermore, if that presumption is to apply to a wide variety of learning algorithms, it must *only* involve the encapsulated information, and not (for example) any characteristics of some class of learning algorithms to which the agent belongs.

For simplicity, consider the case where there are only two possible moves, x^1 and x^2. The encapsulated information provided by n induces a pair of distributions of likely utility values at those two x's, $\int drds\ \delta(\underline{g}_{\rho,s}(x^1, r) - y)\ P(r, s \mid n)$ and $\int drds\ \delta(\underline{g}_{\rho,s}(x^2, r) - y)\ P(r, s \mid n)$, which we can write in shorthand as $P(y; \underline{\gamma}_\rho; n, x^1)$ and $P(y; \underline{\gamma}_\rho; n, x^2)$, respectively. (Note that unlike n, the x^i value in this semicolon notation is a parameter to the random variable $\underline{\gamma}_\rho$, not a conditioning event for that random variable.) By definition of Von Neumann utility functions, for worldview n, the optimal move is x^1 if the expected value $E(y; \underline{\gamma}_\rho; n, x^1) > E(y; \underline{\gamma}_\rho; n, x^2)$, and x^2 otherwise. In general though the learning algorithm of the agent will not (and often cannot) have its distribution over x set to a delta function this way. Other aspects of $P(y; \underline{\gamma}_\rho; n, x^1)$ and $P(y; \underline{\gamma}_\rho; n, x^2)$ besides the difference in their first moments will affect how $P(x \mid n)$ changes in going from the one n to the other. For example, it may be that if $E(y; \underline{\gamma}_\rho; n, x^1) > E(y; \underline{\gamma}_\rho; n, x^2)$, then if n is changed so that both the probability of a relatively large y value at x^2 and the probability of a relatively small y value at x^1 shrinks, while the first moments of those distributions are unchanged, then the algorithm is more likely to choose x^1 with the new n than with the original one.

In light of this, we want to err on the side of caution in presuming how changes to $P(y; \underline{\gamma}_\rho; n, x^1)$ and $P(y; \underline{\gamma}_\rho; n, x^2)$ induced by changing n affect the associated distribution $P(x \mid n)$. The most unrestrictive such presumption we can make is that if the *entire distributions* $P(y; \underline{\gamma}_\rho; n, x^1)$ and $P(y; \underline{\gamma}_\rho; n, x^2)$ are "further separated" from one another after the change in n, then $P(x \mid n)$ gets weighted more to the higher of those two distributions. Such a presumption is the most conservative one we can make that holds for any learning algorithm, i.e., that is cast purely in terms of the set of posterior distributions $\{P(y; \underline{\gamma}_\rho; n, x)\}$ without any reference to attributes of the learning algorithm. This can be viewed as a first-principles justification that it applies to any learning algorithm not horribly mis-suited to the learning problem at hand.[24]

To formalize the foregoing, consider the quantity

$$P(\underline{g}^1 = y^1, \underline{g}^2 = y^2; n, x^1, x^2) \equiv P(\underline{g}_\sigma(x^1, \rho) = y^1 \mid n) P(\underline{g}_\sigma(x^2, \rho) = y^2 \mid n),$$

which expands into the distribution

$$\int \mathrm{d}r^1 \, \mathrm{d}r^2 \, \mathrm{d}s^1 \, \mathrm{d}s^2 \, \delta(\underline{g}_{s^1}(x^1, r^1) - y^1) \delta(\underline{g}_{s^2}(x^2, r^2) - y^2) P(r^1, s^1 \mid n) P(r^2, s^2 \mid n).$$

This is the distribution generated by sampling $P(r', s' \mid n)$ to get values of $\underline{\gamma}_\rho$ at x^1, and then doing this again (in an IID manner) to get values at x^2. This "semicolon" distribution is the most accurate possible distribution of private utilities values at x^1 and x^2 that the agent could possibly employ to decide which x to adopt to optimize that private utility, based solely on n.

Now also fix a utility U that is a single-valued function of x. Our "most accurate distribution" induces the convolution distribution $P(y = y^1 - y^2; n, x^1, x^2)$. The more weighted this convolution is towards values of y that are large and that have the same sign as $U(x^1) - U(x^2)$, the less likely we expect the agent to be "led astray, as far as $U(.)$ is concerned" in "deciding between x^1 and x^2", when the worldview is n. On the other hand, if the convolution distribution is heavily weighted around the value 0, then we expect the agent is more likely to be mistaken (again, as far as U is concerned) in its choice of x.

So consider changing n^a to n^b in such a way that the associated convolution distribution, $P([\underline{g}^1 - \underline{g}^2] \, \mathrm{sgn}[U(x^1) - U(x^2)]; n^a, x^1, x^2)$ is more weighted upwards than is $P([\underline{g}^1 - \underline{g}^2] \, \mathrm{sgn}[U(x^1) - U(x^2)]; n^b, x^1, x^2)$. Say this is the case for *all* pairs of x values (x^1, x^2), i.e., with worldview n^a, the agent is less likely to be led astray for all decisions between a pair of x values than it is with worldview n^b.

[24] If the learning algorithm and underlying distribution over utility values do not adhere to this presumption, then in essence that underlying distribution is "adversarially chosen" for the learning algorithm — that algorithm's implicit assumptions concerning the learning problem are such a poor match to the actual ones — that the algorithm is likely to perform badly for that underlying distribution *no matter what one does* to s, n, or the like.

Our assumption is that whenever such a situation arises, if we truly have an adaptive agent operating in a learnable environment, then the agent has higher intelligence with respect to U, on average, with worldview n^a.

Now in general we can encapsulate how much a stochastic process over C weights some random variable V upward, given some coordinate value $l \in \lambda$, with $\text{CDF}_V(y \mid l)$ — the smaller this cumulative distribution function, the larger the l-conditioned values of V tend to be.[25] Accordingly, we can use such a CDF to quantify how much more "weighted upward" our convolution distribution for n^a is in comparison to the one for n^b. (See App. A for how this CDF is related to intelligence.)

To formalize this we extend the semicolon notation introduced above. Given a coordinate χ whose value c is a single-valued function of (x, r, s), and arbitrary coordinate λ, define the (x^1, x^2, l)-parameterized distribution over values c^1, c^2,

$$\begin{aligned} P(\chi^1, \chi^2; l, x^1, x^2) &\equiv P_\chi(c^1, c^2; l, x^1, x^2) \\ &= \int \mathrm{d}r^1\, \mathrm{d}r^2\, \mathrm{d}s^1\, \mathrm{d}s^2\, P(r^1, s^1 \mid l) P(r^2, s^2 \mid l) \\ &\qquad \delta(\chi(x^1, r^1, s^1) - c^1)\, \delta(\chi(x^2, r^2, s^2) - c^2). \end{aligned}$$

So in this expression χ is a random variable that is (being treated as) parameterized by x, and we are considering its l-conditioned distributions at x^1 and x^2. This notation is sometimes simplified when the meaning is clear, e.g., $P_\chi(c^1, c^2; l, x^1, x^2)$ is written as $P(c^1, c^2; l, x^1, x^2)$.

Expectations, variances, marginalizations, and CDF's of this distribution and of functionals of it are written with the obvious notation. In particular, $P_\chi(c; l, x) = P(\chi(x, \rho, \sigma) = c \mid l)$, so $P_\chi(c^1, c^2; l, x^1, x^2) = P_\chi(c^1; l, x^1) P_\chi(c^2; l, x^2)$. As another example, say that χ is the real-valued coordinate ψ taking values y^i at (x^i, r^i, s^i). Then for any function $f : \Re^2 \rightarrow \Re$, for any l,

$$\begin{aligned} \text{CDF}_{f(y^1, y^2)}(y; l, x^1, x^2) &\equiv \int_{-\infty}^{\infty} \mathrm{d}y^1\, \mathrm{d}y^2\, P(y^1, y^2; l, x^1, x^2) \Theta[y - f(y^1, y^2)] \\ &= \int \mathrm{d}r^1\, \mathrm{d}r^2\, \mathrm{d}s^1\, \mathrm{d}s^2 P(r^1, s^1 \mid l) P(r^2, s^2 \mid l) \\ &\qquad \Theta[y - f(\psi(x^1, r^1, s^1), \psi(x^2, r^2, s^2))]. \end{aligned}$$

Using this notation, for any single-valued function $U : x \rightarrow \Re$, we define the **(ordered) ambiguity** of U and ψ, for l, x^1, x^2, as the CDF of the associated convolution distribution:

$$A(y; U, \psi; l, x^1, x^2) \equiv \text{CDF}_{(y^1 - y^2)\,\text{sgn}[U(x^1) - U(x^2)]}(y; l, x^1, x^2) .$$

Note that the argument of the sgn is just a constant as far as the integrations giving the CDF are concerned. That sgn term provides an ordering of the x's;

[25] Let $\bar{u}$ be a real-valued random variable, and $F : \Re \rightarrow \Re$ a function such that $F(y) > y$; $\forall y \in \Re$. Then $P(F(\bar{u}) < y) \leq P(\bar{u} < y)\ \forall y$, i.e., the monotonically increasing function F applied to the underlying random variable pushes the CDF down. Conversely, if $\text{CDF}_1 < \text{CDF}_2$, then the function $F(u) = \text{CDF}_1^{-1}(\text{CDF}_2(u))$ is a monotonically increasing function that transforms CDF_1 into CDF_2.

ordered ambiguity says how separated our two y-distributions are "in the direction" given by that ordering. When U is not specified, the random variable in the CDF is understood to be $(\psi^1 - \psi^2)$ rather than $(\psi^1 - \psi^2)\,\mathrm{sgn}[U(x^1) - U(x^2)]$. It is easy to verify that such **unordered ambiguities** are related to ordered ones by

$$A(y; U, \psi; l, x^1, x^2) = 1/2 + t_U(x^1, x^2)[A(t_U(x^1, x^2)y; \psi; l, x^1, x^2) - 1/2]$$

where $t_U(x^1, x^2) \equiv \mathrm{sgn}[U(x^1) - U(x^2)]$.

We write just $A(U, \psi; l, x^1, x^2)$ (or $A(\psi; l, x^1, x^2)$) when we want to refer to the entire function over all y. If that entire function shrinks as we go from one n to another — if its value decreases for every value of the argument y — then intuitively, the function has been "pushed" towards more positive values of y. Taking $\lambda = \nu$, such a change will serve as our formalization of the concept that the distributions over U at x^1 and x^2 are "more separated" after that change in the value of ν.

Expanding it in full we can write $A(y; U, \psi; n, x^1, x^2)$ as

$$\int \mathrm{d}r^1\, \mathrm{d}r^2\, \mathrm{d}s^1\, \mathrm{d}s^2\, P(r^1, s^1 \mid l) P(r^2, s^2 \mid l)$$
$$\Theta[y - (\psi(x^1, r^1, s^1) - \psi(x^2, r^2, s^2))\,\mathrm{sgn}[U(x^1) - U(x^2)]],$$

or, by changing coordinates, as

$$\int \mathrm{d}y^1\, \mathrm{d}y^2\, P_\psi(y^1; l, x^1) P_\psi(y^2; l, x^2) \Theta[y - (y^1 - y^2)\,\mathrm{sgn}[U(x^1) - U(x^2)]],$$

and similarly for unordered ambiguities. So ambiguity is parameterized by the two distributions $P(\psi; l, x^i)$ as well as (for ordered ambiguities) U.[26] As a final comment, it is worth noting that there is an alternative to A, A^*, that also reflects the entire n-conditioned CDF of differences in utility values. It and our choice of A rather than A^* is discussed in App. G.

(iii) The first premise

By considering ambiguity with $\psi = \underline{\gamma}_\rho$ and $\lambda = \nu$, we can formalize our the conclusion of reasoning about how certain changes in n affect the probability of the agent's "choosing" a particular x. We call this the **first premise**"

$$\begin{array}{rcl} A(U, \underline{\gamma}_\rho; n^a, x^1, x^2) & < & A(U, \underline{\gamma}_\rho; n^b, x^1, x^2)\ \forall\, x^1, x^2 \\ & \Rightarrow & \\ \mathrm{CDF}(U \mid n^a) & \leq & \mathrm{CDF}(U \mid n^b), \end{array}$$

[26] Note that the ordered ambiguity does not change if we interchange x^1 and x^2, unlike the unordered ambiguity. Note also that unless $\mathrm{sgn}[\psi(x^1, r^1, s^1) - \psi(x^2, r^2, s^2)]$ is the same $\forall\ (r^1, s^1), (r^2, s^2) \in \mathrm{supp}\, P(.,. \mid n)$, the associated ordered ambiguity is non-zero for some $y < 0$. More generally, to have the ambiguity be strongly weighted towards positive values of y, we need that sgn to be the same for all (r', s') in a set with measure (according to $P(r', s' \mid n)$) close to 1.

where U, n^a, and n^b are arbitrary (up to the usual restrictions, that $z \in C$, that U is a function of x, etc.)[27] In other words, we presume that when the condition in the first premise holds, the distribution $P(x \mid n^a)$ must be so much better "aligned" with $U(x)$ than $P(x \mid n^b)$ is that the implication in the first premise (concerning the two associated CDF's) holds. Note that that implication does not involve a specification of r; since in general the agent knows nothing about r, the first premise, which purely concerns $P(x \mid n)$, cannot concern r.

Summarizing, U determines which of the two possible moves x^1 and x^2 by agent ρ are better; $\underline{g}_{\rho,s}$ is the (s-parameterized) private utility that agent ρ is trying to maximize, based exclusively on the value of the worldview, n (a worldview that may or may not provide the agent with the functional form of that private utility

The first premise is, at root, the following assumption: If every one of the ambiguities $A(\underline{\gamma}_\rho; n^a, x^1, x^2)$ (one for each (x^1, x^2) pair) is superior (as far as U is concerned) to the corresponding $A(\underline{\gamma}_\rho; n^b, x^1, x^2)$, then if we replace n^b with n^a, the effect on $P(x \mid n)$ due to that superiority dominates any other characteristics of the two n's. In addition, that dominating effect pushes $P(x \mid n)$ to favor x's having high values of U. As argued above, this is most broadly applicable rule relating certain changes to n and associated changes to an agent's choice of x. There is no alternative we could formulate that is more conservative, i.e., that applies to more learning algorithms, while only involving the distributions of the problem at hand confronting the algorithm.

To explicitly relate the first premise to intelligence, we start with the following result, which has nothing to do with learning algorithms, and which in particular holds regardless of the validity of the first premise. (Indeed, it can be seen as motivating the use of a CDF like ambiguity to analyze properties of intelligences.)

Theorem 3 *Given any coordinates ω, κ and λ, fixed $k \in \kappa$, and two functions $V^a : (w, k) \rightarrow \Re$ and $V^b : (w, k) \rightarrow \Re$ that are mutually factored for coordinate κ,*

$$\begin{array}{rcl} \mathrm{CDF}(V^a \mid l^a, k) & < & \mathrm{CDF}(V^b \mid l^b, k) \\ & \Rightarrow & \\ E(N_{\kappa, V^a} \mid l^a, k) & > & E(N_{\kappa, V^b} \mid l^b, k) \end{array}$$

and similarly when the inequalities are both replaced by equalities.

Now take $\omega = \xi$ and for a fixed k, define $U(.) \equiv V(., k)$ (so that U is a function of x). Then since $P(x \mid n, k) = P(x \mid n)$ (by definition of worldviews), assuming both $P(n^a, k)$ and $P(n^b, k)$ are nonzero, $\mathrm{CDF}(U \mid n^a) < \mathrm{CDF}(U \mid n^b) \Rightarrow \mathrm{CDF}(U \mid n^a, k) < \mathrm{CDF}(U \mid n^b, k) \Rightarrow \mathrm{CDF}(V \mid n^a, k) < \mathrm{CDF}(V \mid n^b, k)$. So if we choose $\lambda = \nu$ in Thm. 3 and combine it with the first premise, we get

[27] Note that the functional (sic) inequality in the first premise is equivalent to $t_U(x^1, x^2) A(\underline{\gamma}_\rho; n^a, x^1, x^2) < t_U(x^1, x^2) A(\underline{\gamma}_\rho; n^b, x^1, x^2)$. In turn, this inequality implies that $U(x^1) \neq U(x^2)$, since otherwise $t_U(x^1, x^2) = 0$.

the promised relation between ambiguities based on the x-ordering $V(x, k)$ and expected κ-intelligences of V conditioned on k and n. In turn, to relate the first premise to the problem of choosing s, use the fact that $E(N_{\kappa,V} \mid n, k, s) = E(N_{\kappa,V}(\xi, \kappa) \mid n, k, s) = E(N_{\kappa,V} \mid n, k)$ to derive the equality $E(N_{\kappa,V} \mid s) = \int dndk P(n, k \mid s) E(N_{\kappa,V} \mid n, k)$.

(iv) Recasting the first premise

Below we will need to use a more general formulation of the first premise than that given above. To derive this more general form, start by defining a parameterized distribution H whose parameter has redundant variables:

$$P(x \mid n) \equiv H_{\{A(\underline{\gamma}_\rho; n, x^1, x^2) : x^1, x^2 \in \xi\}, n}(x)$$

Note that unordered ambiguity is used in this definition, and that H implicitly carries an index identifying the agent as ρ.

In general, the complexity of $P(x \mid n)$ can be daunting, especially if ν is fine-grained enough to capture many different kinds of data that one might have the learning algorithm exploit. This complexity can make it essentially impossible to work with $P(x \mid n)$ directly. However in many situations it is reasonable to suppose that the dependence of H on its ν argument is small in comparison to associated changes in the ambiguity arguments (e.g., n's value does not set a priori biases of the learning algorithm across ξ, etc.). In such situations all aspects of $P(x \mid n)$ get reduced to the dependence of H on ambiguities. In other words, in such situations the functional dependence of $P(x \mid n)$ on the set of ambiguities can be seen as a low-dimensional parameterization of the set of all reasonable learning algorithms $P(x \mid n)$. Accordingly, in these situations one can work with the ambiguities, and thereby circumvent the difficulties with working with $P(x \mid n)$ directly.

Another advantage of reducing $P(x \mid n)$ to H is that often extremely general information concerning $P(\underline{\gamma}_\rho \mid n)$ allows us to identify ways to improve ambiguities, and therefore (by the first premise) improve intelligence. Reduction to H, with its explicit dependence on those ambiguities, facilitates the associated analysis.

In particular, say that the worldview coordinate value specifies the private utility (or at least that we can assume that augmenting the worldview to contain that information would not appreciably change $P(x \mid n)$). This means that $P(\underline{\gamma}_\rho \mid n)$, which arises in calculating ambiguities, can be replaced by $P(\underline{g}_{\rho,s} \mid n)$, where $\underline{g}_{\rho,s}$ is the private utility specified by n. Say that in addition, $P(x \mid n)$ not only is dominated by the the set of associated ambiguities (one ambiguity for each x pair), but can be written as a function exclusively of those ambiguities, a function whose domain is the set of all possible ambiguities. Under these two conditions we could consider the effects on $P(x \mid n)$ of replacing the actual ambiguities $\{A(\underline{\gamma}_\rho; n, x^i, x^j) : x^i, x^j \in \xi\} = \{A(\underline{g}_{\rho,s}; n, x^i, x^j) : x^i, x^j \in \xi\}$, with counterfactual ambiguities $\{A(\underline{g}_{\rho,s'}; n, x^i, x^j) : x^i, x^j \in \xi\}$ that are based on the actual n at hand but are evaluated for some alternative candidate private utility

$\underline{g}_{\rho,s'}$. Under certain circumstances, this approach could be used to determine what such candidate private utility to use, based on comparing the associated counterfactual ambiguities.

To use this approach in as broad a set of circumstances as possible, we must address the fact that $P(x \mid n)$ may have some dependence on n not fully captured in the associated ambiguities, e.g., when n modifies the learning algorithm, for example by specifying biases for the learning algorithm to use. This means the definition given above for H will not in general extend to parameter values whose ambiguity set does not correspond to n. Another hurdle is that often the domain of $P(x \mid n)$ need not extend to all ambiguities of the form $\{A(\underline{g}_{\rho,s'}; n, x^i, x^j) : x^i, x^j \in \xi\}$. Finally, in general worldviews do not specify the private utility.

To circumvent these difficulties we need to introduce new notation and recast the first premise accordingly. Start by extending the domain of definition of H to write it as $H_{\{A(\psi;l,x^1,x^2):x^1,x^2 \in \xi\},n}(x)$, for any coordinate value $l \in \lambda \subseteq \nu$. Here ψ is an arbitrary real-valued function of x, r, and s, not necessarily related to $\underline{\gamma}_\rho$. So $H_{\{A(\psi;l,x^1,x^2):x^1,x^2 \in \xi\},n}(x)$ is not necessarily related to the actual $P(x \mid n)$. Despite these freedoms, we require that for any value of its parameters $H_{\{A(\psi;l,x^1,x^2):x^1,x^2 \in \xi\},n}(x)$ is a proper probability distribution over x, one that for fixed ψ and $\lambda = \nu$ is (like $P(x \mid n)$) parameterized by n. This extending of H's domain is how we circumvent the first two of our difficulties.

Next we introduce some succinct notation. As in the definition of worldviews let $W \in \Omega$ refer to the set of all non-ξ coordinate we will consider in our analysis, and define the distribution $P^{[\psi;\lambda]}(x, l, W) \equiv H_{\{A(\psi;l,x^1,x^2):x^1,x^2 \in \xi\},n}(x) P(l, W)$, where $\lambda \subseteq \nu$. When $\psi = \underline{\gamma}_\rho$, we just write $P^{[\lambda]}$. So for example $P^{[\nu]}(x \mid n) = P^{[\underline{\gamma}_\rho;\nu]}(x \mid n) = P(x \mid n)$, $P^{[\psi;\lambda]}(x \mid l, W) = P^{[\psi;\lambda]}(x \mid l) = H_{\{A(\psi;l,x^1,x^2):x^1,x^2 \in \xi\},n}(x)$, etc. Note also that $P^{[\underline{\gamma}_\rho;\nu,\sigma]}(x \mid n, s) = P^{[\underline{g}_{\rho,s};\nu,\sigma]}(x \mid n, s)$. Intuitively, we view the learning algorithm as taking arbitrary sets ambiguities and worldviews as input and producing a distribution over x; $P^{[\psi;\lambda]}(x \mid l)$ is the distribution over x that arises when the learning algorithm is fed the ambiguities $\{A(\psi; l, x^1, x^2) : x^1, x^2 \in \xi\}$ and worldview n specified by l.

Now consider the following elementary result:

Lemma 1 *Consider any two probability density functions over the reals, P_1 and P_2, where $\frac{P_1(u)}{P_1(u')} \geq \frac{P_2(u)}{P_2(u')}$ $\forall u, u' \in \Re$ where $u > u'$. Say we also have any $\phi : \Re \rightarrow \Re$ with nowhere negative derivative. Then $CDF_{P_1}(\phi) \leq CDF_{P_2}(\phi)$.*

Combining this lemma with the first premise, and using our new notation, we arrive at the following version of the first premise, derived in the appendix:

Theorem 4 *Given coordinate values l^a and $l^b \in \lambda \subseteq \nu$, $\exists H$ such that*

$$\begin{aligned} A(U, \psi^a; l^a, x^1, x^2) &< A(U, \psi^b; l^b, x^1, x^2) \;\; \forall x^1, x^2 \\ &\Rightarrow \\ \mathrm{CDF}^{[\psi^a;\lambda]}(U \mid l^a) &\leq \mathrm{CDF}^{[\psi^b;\lambda]}(U \mid l^b), \end{aligned}$$

where as usual ψ^a, ψ^b and (the r-independent) U are arbitrary.

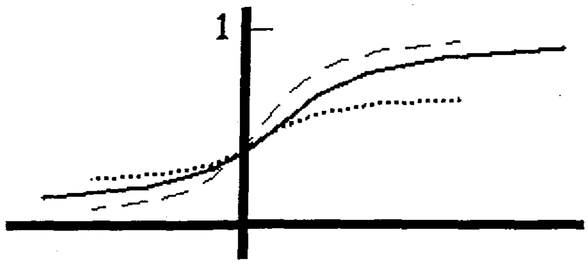

Figure 1: The solid line depicts an ambiguity $A(y; V; l, x^1, x^2)$. The dotted line depicts $A(y; KV; l, x^1, x^2) = A(y/K; V; l, x^1, x^2)$ for $K > 1$; the dashed line is $A(KV; l, x^1, x^2)$ for $0 < K < 1$. Neither of those scaled-utility ambiguities lies entirely below the original one. Accordingly, neither of those scaled utilities is recommended by the first premise.

This theorem is illustrated geometrically in Fig. 1.

Because it holds for any underlying distribution over ζ, Thm. 3 holds for CDF's and expectation values based on any $P^{[\psi;\lambda]}$, not just $P^{[\underline{\gamma}_\rho;\nu]}$. Since for any ψ, $P^{[\psi;\lambda]}(x \mid l, W) = P^{[\psi;\lambda]}(x \mid l)$, the discussion following Thm. 3 holds for $P^{[\psi;\lambda]}$ conditioned on l just as well as for P conditioned on n. So Thm. 4 has the following corollary:

Corollary 1 *Given any coordinates κ and $\lambda \subseteq \nu$, fixed $k \in \kappa$, and $V : (x, k) \rightarrow \Re$, $\exists H$ such that*

$$\begin{array}{ccc} A(V(.,k), \psi^a; l^a, x^1, x^2) & < & A(V(.,k), \psi^b; l^b, x^1, x^2) \;\; \forall\, x^1, x^2 \\ & \Rightarrow & \\ E^{[\psi^a;\lambda]}(N_{\kappa,V} \mid l^a, k) & \geq & E^{[\psi^b;\lambda]}(N_{\kappa,V} \mid l^b, k) \end{array}$$

Summarizing, for a particular value of k, V determines which of the two possible moves x^1 and x^2 by agent ρ are better; $\underline{g}_{\rho,s}$ is the (s-parameterized) private utility that agent ρ is trying to maximize, based exclusively on the value of the worldview, n (a worldview that may or may not provide the agent with the functional form of that private utility); ψ^a and ψ^b are two real-valued functions of x, r and s that are used to evaluate ambiguities, and l^a and l^b are values of a conditioning variable for evaluating ambiguities, a variable that specifies n at a minimum. In addition, H is a parametrized distribution over x that is defined for any parameter value that consists of $O(\xi)$ CDF's and a worldview, a distribution that equals $P(x \mid n)$ when the its parameter value is the set $\{A(\underline{\gamma}_\rho; n)\}$ together with n, and more generally for any $\lambda \subseteq \nu$ is expressed as $P^{[\psi;\lambda]}(x \mid l)$ whenever the CDF's are the ambiguities $\{A(\psi; l, x^1, x^2) : x^1, x^2 \in \xi)\}$. From now on, unless explicitly stated otherwise, we will assume that we are restricting attention to an H for which Coroll. 1 holds.

(v) The second premise

Having rewritten the first premise this way, we can address the potential problem arising when the worldview does not specify the private utility. First consider any changes to s that modify the associated set of n for which $P(n \mid s)$ is substantial. Typically, any such change in the likely n fixes fairly precisely what the inducing changes in s are, as far as evaluation of ambiguities is concerned. Accordingly, when exploiting the first premise we usually restrict attention to scenarios in which $\forall r \in \operatorname{supp} P(r \mid s)$ we can approximate

$$\begin{aligned}&\int \mathrm{d}n P(n \mid s) P^{[\nu]}(x \mid n) \quad = \\ &\int \mathrm{d}n P(n \mid s) P^{[\nu,\sigma]}(x \mid n, s).\end{aligned}$$

We refer to this approximation as the **second premise**. Note that it holds exactly if n contains a specification of $\underline{g}_{\rho,s}$, and $P(x \mid n)$ only depends on the associated ambiguities, $\{A(\underline{\gamma}_\rho; n, x^i, x^j)\} = \{A(\underline{g}_{\rho,s}; n, x^i, x^j)\}$. So if we can treat the system as though this were the case, on average, then the second premise holds.[28] A semi-formal example of a more general situation where the second premise holds is presented in App. F.[29]

The following corollary of the second premise is often useful:

Corollary 2 *Where V is any utility function, $h \in \eta$ any coordinate, and $W \in \Omega$ any non-ξ coordinate,*

$$E(V \mid h, s) = \int \mathrm{d}n\mathrm{d}W \; P(W \mid s) P(n \mid W, s) \; E^{[\underline{g}_{\rho,s};\nu,\sigma]}(V \mid n, s, h, W)$$

Often this result can be used in conjunction with Coroll. 1 to analyze the implications of various choices of s. As an example, in many situations (e.g., in very large systems) changes to ρ's private utility will have relatively little effect on the rest of the system, i.e., will have minimal effect on the distribution over r values. Accordingly consider s^a and s^b that vary only in that choice of ρ's private utility[30], in a situation where this implies that $P(r \mid s^a) = P(r \mid s^b) \equiv P(r \mid s^{ab})$.

[28] Conversely, if σ is "perniciously chosen" to always force n to equal n' for any s, where n' gives no information about the likely values that s is inducing of $\underline{g}_{\rho,s}$ at the various r, then $\int \mathrm{d}n P(n \mid s) P^{[\nu]}(x \mid n) = P(x \mid n')$ and does not reflect the ambiguities determining $\int \mathrm{d}n P(n \mid s) P^{[\nu,\sigma]}(x \mid n, s) = P^{[\nu,\sigma]}(x \mid n', s)$. In such a situation the second premise will not hold. This is similar to the situation with the first premise; in both an adversarially poor match between the learning algorithm and the learning problem at hand confounds our premise.

[29] If it weren't for the second premise, we would have to work with $P(r \mid n)$ rather than $P(r \mid n, s)$ in evaluating ambiguities. This would then require specifying a prior $P(\bar{s})$, reflecting "prior beliefs" of what the private utility is likely to be, among other aspects of s. Specifying a prior over such a space and then integrating against it can be a fraught exercise. In essence, the second premise allows us to circumvent this when averaging over n, by setting that prior to a delta function about the actual s. Nonetheless, it is important to note that we do not need a hypothesis as powerful as the second premise to do this; the second premise is only used once, in the proof of Coroll. 3 below, and a significantly weaker version of it would suffice there. We present the "powerful" version instead for pedagogical clarity.

[30] Formally, our presumption is that $\forall \; z^a \in s^a$, $z^b \in s^b$, $\sigma\text{-}_{\underline{g}}(z^a) = \sigma\text{-}_{\underline{g}}(z^b)$.

Let V be a utility function, so that $N_{\rho,V}$ is as well. Then for both $s = s^a$ and $s = s^b$, by using Coroll. 2 with $\Omega = \rho$ and $\eta = \emptyset$, we establish that

$$E(N_{\rho,V} \mid s) \;\; = \;\; \int \mathrm{d}r\mathrm{d}n P(r \mid s^{ab}) P(n \mid r, s) \; E^{[\underline{g}_s;\nu,\sigma]}(N_{\rho,V} \mid n, r, s).$$

So by Coroll. 1, taking $\lambda = \nu \cap \sigma$, $\kappa = \rho$, and $\psi^a = \psi^b = \underline{\gamma}_\rho$, if separately for each r for which $P(r \mid s^{ab})$ is substantial,

$$A(V(.,r), \underline{\gamma}_\rho; n^a, s^a, x^1, x^2) < A(V(.,r), \underline{\gamma}_\rho; n^b, s^b, x^1, x^2) \;,$$

(for all (x^1, x^2) pairs, and for all (n^a, n^b) such that both $P(n^a \mid r, s^a)$ and $P(n^b \mid r, s^b)$ are substantial) we can conclude that $E(N_{\rho,V} \mid s^a) > E(N_{\rho,V} \mid s^b)$. This approach can be used even if the coordinate utility V is factored with respect to G but the private utility is not. Note also that if we take $V = \underline{g}_{\rho,s^b}$ and have $\underline{g}_{\rho,s^a}$ be factored with respect to $\underline{g}_{\rho,s^b}$, then our reasoning implies that $E(N_{\rho,\underline{g}_{\rho,s^a}} \mid s^a) > E(N_{\rho,\underline{g}_{\rho,s^b}} \mid s^b)$.

The first two premises can also be used to analyze the effect on agent ρ of changes to the *other* agents. In addition they can be used to analyze changes that amount to a complete redefinition of the agent (which changes we can implement by inserting commands in the value of the agent's worldview that change how it behaves), or more generally, a coordinate transformation [22]. Indeed, by those premises, H, $\underline{\gamma}_\rho$ and $P(r \mid n, s)$ parameterize $P(x \mid n)$. In particular, say $\sigma = \sigma_{\underline{\gamma}_\rho} \subseteq \nu$, H has no direct dependence on n not arising in the ambiguities, and we take $P(r \mid s)$ to be uniform. Then for fixed H, all aspects of the learning algorithm are set by $\underline{\gamma}_\rho$, $P(n \mid r, s)$, and the associated ambiguities.

More generally, once we specify $P(r \mid s)$ in addition to these quantities, we have made all the choices available to us as designers that affect term 3 of the central equation. In principle, this allows us to solve for the optimal one of those four quantities given the others. For example, for fixed $\underline{\gamma}_\rho$, H, and $P(r \mid s)$, we could solve for which $P(n \mid r, s)$ out of a class of candidate such likelihoods optimizes expected intelligence.[31]

The rest of this paper presents a few preliminary examples of such an approach, concentrating on changes to s that only alter one or more agents' private utilities, where only very broad assumptions about $P(n \mid r, s)$ are used. These are the scenarios in which the premises have been most thoroughly investigated, and therefore in which confidence that H etc. do indeed capture the totality of a learning algorithm is highest.

(vi) The third premise

As just illustrated, for some differences in s (namely those that only modify private utilities), we can simplify the analysis to involve only a single s-induced

[31] More formally, where $\sigma \subseteq \sigma_\nu$ sets the likelihood $P(n \mid r, s_r ho, s_\nu)$, we could solve for the s_ν optimizing expected intelligence.

distribution over r's (namely $P(r \mid s^{ab})$). The analysis still involved different distributions over n's however, one each of the two s's (in the guise of the two distributions $P(n \mid r, s)$). Moreover, to calculate expected intelligence for a given s we must average over n, and usually changes to s change $P(n \mid r, s)$ in a way difficult to predict.[32] Therefore to exploit the first two premises to determine which of the two s's gave better expected intelligence, we had to have a desired difference in ambiguities hold for *all* pairs of n's generated from the two s's, an extremely restrictive condition.

One way around this would be to extend the analysis in a way that only involves a single s-induced distribution over n's. To see how we might do this, fix r, x^1, and x^2, and consider a pair s^a and s^b that differ only in the associated private utility for agent ρ, where those two utilities are mutually factored. Train on $\underline{g}_{s^b}$, thereby generating an n according to $P(n \mid r, s^b)$, and thence a distribution over r', $P(r' \mid n)$, which in turn gives an ambiguity between values of the private utility at x^1 and x^2 and therefore an expected intelligence. Our choice of private utility affects this process in three ways:

1) By affecting the likely n, and therefore $P(r' \mid n)$.

2) By affecting how well distinguished utility values at x^1 and x^2 are for any associated pair of r' values generated from $P(r' \mid n)$. If $P(r' \mid n)$ is broad and/or the private utility is poor at distinguishing x^1 and x^2, then ambiguity will be poor.

3) By providing one of the arguments to H which (given the utility, and along with the ambiguities of (2)) fixes the distribution over intelligences.

In the guise of Coroll. 1 (with $\lambda = \nu$, $\kappa = \Omega = \rho$, $\psi^a = \underline{g}_{s^a} = V^a$, and $\psi^b = \underline{g}_{s^b} = V^b$), the first premise concerns the second effect. If we combine this with the second premise (in the guise of Coroll. 2, with $\Omega = \rho$), we see that the first two premises concern the last two effects of the choice of private utility on expected intelligence. They say nothing about the first effect of the private utility choice though.

It is typically the case that the first effect will tend to work in a correlated manner with the last two effects. That is, if for some given n generated from $\underline{g}_{\rho,s^b}$ the utility $\underline{g}_{\rho,s^a}$ results in higher intelligences (e.g., because it is better able to distinguish utility values than is $\underline{g}_{\rho,s^b}$), it is typically also the case that if one had used $\underline{g}_{\rho,s^a}$ to generate n's in the first place, it would have resulted in more informative n, and therefore $P(r' \mid n)$ would have been crisper, leading to a better ambiguity and thence expected intelligence.

We formalize this as the **third premise.**[33]

[32] For example, in a multi-stage game (see App. D), in general changing $\underline{g}_{\rho,s}$ causes our agent to take different actions at each stage of the game, which usually then causes the behavior of the other agents at later stages to change, which in turn changes ρ's training data, contained in the value of n at those later stages.

[33] An alternative to the version of the third premise presented here that would serve our purposes just as well would have all distributions conditioned on some $b \in \beta \subseteq \sigma$ (e.g., (r, s)), rather than just on s. One could also modify the hypothesis condition of the third premise by

Say that s^a and s^b differ only in their associated private utilities, and that those utilities are mutually factored. Then

$$\begin{aligned}\int dnP(n \mid s^b)E^{[\underline{g}_{s^a};\nu,\sigma]}(N_\rho \mid n, s^b) &\geq \int dnP(n \mid s^b)E^{[\underline{g}_{s^b};\nu,\sigma]}(N_\rho \mid n, s^b) \\ &\Rightarrow \\ \int dnP(n \mid s^a)E^{[\underline{g}_{s^a};\nu,\sigma]}(N_\rho \mid n, s^a) &\geq \int dnP(n \mid s^b)E^{[\underline{g}_{s^b};\nu,\sigma]}(N_\rho \mid n, s^b).\end{aligned}$$

Together with Coroll. 2 this results in the following:

Corollary 3 *Say s^a and s^b differ only in the associated private utility for agent ρ, and that those utilities are mutually factored. Then*

$$\begin{aligned}\textstyle\int dndrP(r \mid s^b)P(n \mid r, s^b)E^{[\underline{g}_{s^a};\nu,\sigma]}(N_\rho \mid n, r, s^b) \quad &\geq \\ \textstyle\int dndrP(r \mid s^b)P(n \mid r, s^b)E^{[\underline{g}_{s^b};\nu,\sigma]}&(N_\rho \mid n, r, s^b) \\ &\Rightarrow \\ E(N_{\rho,\underline{g}_{s^a}} \mid s^a) \quad \geq \quad &E(N_{\rho,\underline{g}_{s^b}} \mid s^b).\end{aligned}$$

If, $\forall r, A(\underline{g}_{\rho,s^b}(\xi, r), \underline{g}_{\rho,s^b}; n, x^1, x^2, s^b) > A(\underline{g}_{\rho,s^b}(\xi, r), \underline{g}_{\rho,s^a}; n, x^1, x^2, s^b)$ (for all (x^1, x^2), and for all n such that $P(n \mid r, s^b)$ is substantial), then by Coroll. 1 the condition in Coroll. 3 is met (take $\lambda = \nu \cap \sigma$ and $\kappa = \rho$, as usual). So by Coroll. 3, in such a situation we can conclude that $E(N_{\rho,\underline{g}_{s^a}} \mid s^a) \geq E(N_{\rho,\underline{g}_{s^b}} \mid s^b)$, i.e., that for fixed r, s^a has better term 3 of the central equation than does s^b. This is the process that will be the central concern of the rest of this paper: inducing improved ambiguity, and then plugging the first premise (in the guise of Coroll. 1) into the second and third premises (combined in Coroll. 3) to infer improved expected intelligence.

In particular, again consider the situation (discussed in the subsection on the first premise) where $P(r \mid s^a) = P(r \mid s^b) \equiv P(r \mid s^{ab})$, and assume this also equals $P(r \mid s^b)$. If separately for each r for which $P(r \mid s^{ab})$ is substantial, and for all associated n for which $P(n \mid r, s^{ab})$ is substantial,

$$A(\underline{g}_{\rho,s^b}(., r), \underline{g}_{\rho,s^b}; n, x^1, x^2, s^b) \;>\; A(\underline{g}_{\rho,s^b}(., r), \underline{g}_{\rho,s^a}; n, x^1, x^2, s^b),$$

then we can conclude that

$$E(N_{\rho,\underline{g}_{s^a}} \mid s^a) \geq E(N_{\rho,\underline{g}_{s^b}} \mid s^b).$$

replacing s^b throughout with some alternative s^*, and our results would still hold under the substitution throughout of $s^b \rightarrow s^*$. Similarly one could change the integration variable $n \in \nu$ to some other coordinate $l \in \lambda \subseteq \nu$. For all such changes the results presented below — and in particular Coroll. 3 — would still hold; the important thing for those results is that each ambiguity arising in the integrand of the left-hand-side of the hypothesis condition of the third premise is evaluated with the same distribution over r^1 and r^2 as the corresponding ambiguity in the right-hand-side. For pedagogical clarity though, no such modification is considered here.

Of course, in practice this condition won't hold for all such r and n. At the same time, Coroll. 3 makes clear that it doesn't need to; we just need the associated integrals over r and n to favor s^a over s^b.

(vii) Example: The collapsed utility

As an example of how to use Coroll. 3, consider the use of a Boltzmann learning algorithm for our agent [25], where s^b is our original s value. With such an algorithm, constructing a new private utility by scaling the original one (i.e., changing s) is equivalent to modifying the learning algorithm's temperature parameter. Now say that for any pair of moves, the ambiguity for s^b and any probable associated worldview n^b is zero for all negative y values. Then changing s by lowering the temperature will monotonically lower $A(\underline{g}_{\rho,s^b}(\xi, r), \underline{g}_{\rho,s}; n^b, x^1, x^2)$. Accordingly, doing this cannot lower expected intelligence, only increase it. (Note that the new private utility is factored with respect to the original one, so this effect of changing s also holds for expected intelligence with respect to the original private utility.)

Now consider the following theorem:

Theorem 5 *Fix* $n, s^a, s^b, r \in \operatorname{supp} P(. \mid s^b)$ *and a function* $U : x \in \xi \to \Re$. *Stipulate that*

i) $\forall\, x, x' \in \xi, \operatorname{sgn}[U(x,r) - U(x',r)] = \operatorname{sgn}[\underline{g}_{s^b}(x,r) - \underline{g}_{s^b}(x',r)]$;

ii) $\forall\ r' \in \operatorname{supp} P(. \mid n)$, *there exists two real numbers* $A_{r'}$ *and* $B_{r'} \leq A_{r'}$ *such that* $\underline{g}_{s^b}(x, r')$ *takes on both values—but no others—as one varies the* $x \in \xi$;

iii) for all such r' $\underline{g}_{s^a}(x, r') = 0$ *if* $A_{r'} = B_{r'}$, *and equals* $\frac{\underline{g}_{s^b}(x,r') - B_{r'}}{A_{r'} - B_{r'}}$ *otherwise, and* $\forall\ r' \notin \operatorname{supp} P(. \mid n)$, $\underline{g}_{s^a}$ *is factored with respect to* $\underline{g}_{s^b}$*;*

iv) for each pair of moves, for at least one move of that pair, x^*, $\exists\ y^*$ *such that* $P(\underline{g}_{s^a}(x^*, \rho) = y \mid n) = \delta(y - y^*)$.

Then $\forall\ x^1, x^2$, $A(U, \underline{g}_{s^a}; n, x^1, x^2)$ *has purely non-negative support.*

(An analogous version of this result holds if instead we take $\underline{g}_{s^a}(x, r') = 1$ whenever $A_{r'} = B_{r'}$.)

Condition (i) of Thm. 5 can be viewed as a weakened form of requiring that U and $\underline{g}_{s^b}$ be factored. In particular, it trivially holds for $U = \underline{g}_{s^b}$, or (due to the fact that $\underline{g}_{s^a}$ is a difference utility with lead utility $\underline{g}_{s^b}$) $U = \underline{g}_{s^a}$. Conditions (ii) and (iii) mean that for each r', the values of $\underline{g}_{s^a}(x, r')$ as one varies x are those of $\underline{g}_{s^b}$ "collapsed" to one of the two values 0 or 1. However for fixed x, which of that pair of values equals $\underline{g}_{s^a}(x, r')$ can differ from one r' to the next.

There are many situations in which condition (ii) of Thm. 5 holds with $\underline{g}_{s^b} = G$. One example is a spin glass with G given by the Hamiltonian. Another is the simple spin system where $G(z) = \sin(\pi n(z)/2)$, $n(z)$ being defined as the total number of spins in the up configuration.

Condition (iv) means that given worldview n, context r, and a pair of moves, there is no room for uncertainty in the value of the private utility at x^*—it must equal (the typically unknown value) y^* there. (Note that which element of the pair of moves is this special x can vary with n and/or r.) This will often be the case if, for example, n was generated from $\underline{g}_{s^a}$, and the agent's (n-based) "prediction" for the utility value of the particular move it actually ends up making is both unambiguous and correct. In particular, such prediction accuracy often can be induced by having all the *other* agents readily "freeze" into a static background. In turn, as an example, those other agents are likely to freeze if they all use Boltzmann learning algorithms with their temperatures set low enough, and with the windows they use to estimate the utilities of their possible moves short enough.

We call the difference utility $\underline{g}_{s^a}$ in Thm. 5 the **collapsed utility** (CU), and say that it is formed by **collapsing** $\underline{g}_{s^b}$, since for fixed r' it is formed by collapsing all the values $\underline{g}_{s^b}(x, r')$ takes on as one varies x, to either 0 or 1.

When the conditions in Thm. 5 hold the ambiguity will shrink monotonically as the CU is scaled upwards. As an example, consider a Boltzmann learning algorithm in the scenario discussed at the end of the previous subsection, where in addition the conditions in Thm. 5 are met for private utility set to the CU. As the temperature parameter of that algorithm shrinks the associated expected intelligence cannot decrease, and should in particular eventually exceed that of $\underline{g}_{s^b}$.[34] Therefore for the choice of $\underline{g}_{s^b} = G$, the value of G induced by using CU as the private utility with a low enough temperature should be larger than that induced by using the team game at any temperature.

3 The Aristocrat and Wonderful Life Utilities

In this section we illustrate a general set of techniques for changing the private utility so as to monotonically lower unordered ambiguity conditioned on a particular n. As discussed above, when plugged into Coroll. 3 such improved ambiguities can cause the new private utility to have better expected intelligence than the original one.

The analysis will be closely analogous to that behind the use of Fisher's linear discriminant in statistics. We will start by restricting the analysis to distributions obeying a linearity condition. This is essentially an extended form of assuming Gaussian distributions — such an assumption being the starting point of the derivation of Fisher's linear discriminant. We will then exploit Coroll. 3 to derive "learnability" as a measure of the quality of a private utility (as far as term 3 in the central equation is concerned). Formally, learnability

[34] Formally, the fact that ambiguity for $\underline{g}_{s^a}$ has purely non-negative support does not mean that the ambiguity for $\underline{g}_{s^b}$ has a support that extends to negative values. In practice though, that is the case for the vast majority of $n \in \text{supp}P(. \mid s^b)$. Even so, we cannot conclude that the ambiguity function for $\underline{g}_{s^a}$, extending over all y, is less than that for $\underline{g}_{s^b}$. We can conclude that the reverse does not hold though. And again, in practice, the discrepancy in supports usually does mean that the ambiguity function for $\underline{g}_{s^a}$ is less than that for $\underline{g}_{s^b}$, so that we can apply the first two corollaries premises.

is identical to the Rayleigh coefficient, just expressed in a different setting. Completing the analogy, whereas with the Fisher discriminant one strives for coordinate transformations of a data set giving a large value of the associated Rayleigh coefficient, at the end of this section we demonstrate transformations to the private utility giving a large value of the associated learnability.

(i) Learnability

We begin by considering the first order expansion of the distribution of one utility in terms of the distribution of another utility:

Theorem 6 *Fix $l, l' \in \lambda \subseteq \nu, x^1, x^2$, an x-ordering U, and two utilities V_a and V_b, where $\exists\ K \in \Re^+$ and $h : \xi \rightarrow \Re$ such that*

$$P_{V_a}(y^1, y^2; l', x^1, x^2) = P_{KV_b+h}(y^1, y^2; l, x^1, x^2).$$

Then $\forall\ y$,

$$A[y; U, V_a; l', x^1, x^2] = A\left[\frac{y}{K} + t_U(x^1, x^2)\left(\frac{h(x^2) - h(x^1)}{K}\right); U, V_b; l, x^1, x^2\right].$$

So if in addition to the condition in Thm. 6, $\forall\ y$,

$$A\left[\frac{y}{K} + t_U(x^1, x^2)\left(\frac{h(x^2) - h(x^1)}{K}\right); U, V_b; l, x^1, x^2\right] < A[y; U, V_b; l, x^1, x^2],$$

then it follows that $A[U, V_a; l', x^1, x^2] < A[U, V_b; l, x^1, x^2]$.

We will sometimes find it convenient to put subscripts on K and/or h explicitly giving the values of l', V_a, l, V_b, x^1 and/or x^2, in that order. For example, in Fig. 2 we refer to $K_{V,U}$, to mean K when $V_a = V$ and $V_b = U$.[35]

It is often the case that "to first order", changing from $V = V_b$ to $V = V_a$ doesn't change the shapes of any of the associated distribution functions $P(V(x) = v \mid l)$ (one such distribution for each x). Primarily, all the change does to those distributions is separately shift them, and/or contract them all by the same factor.[36,37] The condition in Thm. 6 is (a slightly weaker version

[35] Note the following algebraic rules concerning such sets of distributions that are linearly related:

$$\begin{aligned} K_{l_1,V_1,l_3,V_3} &= K_{l_1,V_1,l_2,V_2}K_{l_2,V_2,l_3,V_3}; \\ K_{l_1,V_1,l_2,V_2} &= 1/K_{l_2,V_2,l_1,V_1}; \\ h_{l_1,V_1,l_3,V_3} &= K_{l_1,V_1,l_2,V_2}h_{l_2,V_2,l_3,V_3} - h_{l_1,V_1,l_2,V_2}; \\ h_{l_1,V_1,l_2,V_2} &= -h_{l_2,V_2,l_1,V_1}/K_{l_2,V_2,l_1,V_1}. \end{aligned}$$

[36] This is particularly common in situations where there are extremely many possible V values, densely packed together.

[37] Note that a linear relationship between utilities is a sufficient but not necessary condition for a linear relationship between the distributions of their values.

of) the requirement that this property holds exactly, even if we also switch from l to l' at the same time (and therefore change the underlying probability distribution over z). The general effects of expansion or contraction of the utility on the associated ambiguity are illustrated in Fig. 2.

Thm. 6 tells us in particular that when its condition is met along with the one mentioned just following its presentation, then for $K = 1$ and $t_U(x^1, x^2)[h(x^2) - h(x^1)]$ negative, then changing from (V_b, l) to (V_a, l') improves ambiguity. Moreover, the degree of that drop grows with increasing magnitude of $\{h(x^2) - h(x^1)\}/K$.[38] In the usual way, for $l = l', \lambda = \nu \cap \sigma, V_a = \underline{g}_{s'}$, and $V_b = \underline{g}_s$, where s and s' only differ in their private utilities, we can exploit this phenomenon in concert with Coroll. 1 and then Coroll. 3 to improve term 3. To that end we start with the following:

Theorem 7 *Say that the condition in Thm. 6 holds for the quadruple* (l', V_a, l, V_b) *with the same* $K, h \; \forall x^1, x^2$. *Then*

i) where f *is any distribution over* x,

$$K = \sqrt{\frac{\int \mathrm{d}x \, f(x) \, \mathrm{Var}(V_a; l', x)}{\int \mathrm{d}x \, f(x) \, \mathrm{Var}(V_b; l, x)}}.$$

Now define

$$\Lambda_f(U; l'', x^1, x^2) \equiv \frac{E(U; l'', x^1) - E(U; l'', x^2)}{\sqrt{E_{f(x)}(\mathrm{Var}(U; l'', \xi))}}$$

where $E_{f(x)}(\mathrm{Var}(U; l'', \xi)) = \int \mathrm{d}x \, f(x) \, \mathrm{Var}(U(x, \rho) \mid l'')$. *Then*

ii)

$$\frac{h(x^2) - h(x^1)}{K} \propto \Lambda_f(V_b; l, x^1, x^2) - \Lambda_f(V_a; l', x^1, x^2),$$

where the V_a*-independent proportionality constant is* $\sqrt{\int \mathrm{d}x \, f(x) \, \mathrm{Var}(V_b; l, x)}$.

We call $\frac{h(x^2) - h(x^1)}{K}$ the **(ambiguity) shift** and $\Lambda_f(U; l, x^1, x^2)$ the **learnability** of U for x^1, x^2, and l.[39] As a particular example, for $f(x) = (1/2)[\delta(x - x^1) + \delta(x - x^2)]$,

$$[\Lambda_f(U; l'', x^1, x^2)]^2 = 2 \frac{[E(U; l'', x^1) - E(U; l'', x^2)]^2}{\mathrm{Var}(U; l'', x^1) + \mathrm{Var}(U; l'', x^2)}. \tag{2}$$

Note that $|\Lambda_f(U; l'', x^1, x^2)|$ is invariant under affine transformations of U. Typically we are interested in the case where $\mathrm{sgn}[E(V_b^1 - V_b^2; l, x^1, x^2)] = \mathrm{sgn}[E(V_a^1 -$

[38] A similar result holds if we instead consider a fixed pair (x^1, x^2) and associated K_{x^1, x^2}, so that the expansion factor can vary with moves, just like the offset factor h.

[39] This latter is a slight modification from the definition used in our previous work.

$V_a^2; l, x^1, x^2)] = t_{V^b}(x^1, x^2)$, so that we can use learnability to evaluate the offset term in Thm. 6, $t_U(x^1, x^2)\left[\frac{h(x^2)-h(x^1)}{K}\right]$.

Intuitively, the learnability of U reflects its signal-to-noise, as far as agent ρ is concerned, in that agent's process of "choosing its move". This is because the numerator term in the definition of learnability reflects how much (the expectation of) that utility varies as one changes the agent's move x with the context held fixed. In contrast, the denominator term reflects the (average over x of) how much U varies due to uncertainty in the context while keeping the move x fixed.[40]

The following results provide a geometric perspective on the expressions in Thm. 7:

Theorem 8 *Say that the condition in Thm. 6 holds for the quadruple (l', V_a, l, V_b).*

i) If both V_a and V_b are difference utilities with the same lead utility and $\beta = 1$, while both $P(r'; l) = P(r'; l')$ and $\Lambda_f(V_b; l, x^1, x^2) < \Lambda_f(V_a; l', x^1, x^2)$, then $K < 1$.

ii) Let $\{V_a, l'\}$ be an equivalence class of (V, l) pairs all related to (V_b, l) as in Thm. 6. Then the learnability of those pairs multiplied by $t_U(x^1, x^2)$ is a shrinking function of the value of the associated ambiguities at the origin. In addition, across all pairs in that class that share some particular learnability value, K is inversely proportional to the slope of the ambiguity of that pair at the origin.

iii) Say the condition (ii) also holds for the quadruple $(l^*, V_{a^*} \equiv \beta V_a, l, V_b)$ (though potentially for a different K and/or h), where $P(r'; l^*) = P(r'; l')$. Then $\Lambda_f(V_{a^*}; l^*, x^1, x^2)$ and $\Lambda_f(V_a; l', x^1, x^2)$ are identical $\forall\, x^1$, x^2, as are the associated shifts, while $K_{l^*, V_{a^*}, l, V_b} = \beta K_{l', V_a, l, V_b}$.[41]

iv) If $K < 1$ and $\Lambda_f(V_a; l', x^1, x^2) > \Lambda_f(V_b; l, x^1, x^2)$ ($K > 1$ and $\Lambda_f(V_a; l', x^1, x^2) < \Lambda_f(V_b; l, x^1, x^2)$, respectively), then the maximal slope of $A(V_a; l', x^1, x)$ is greater than (less than, respectively) the maximal slope of $A(V_b; l, x^1, x^2)$.

To understand Thm. 7 in terms of ambiguities, for pedagogical simplicity consider making changes to a utility V without any corresponding changes to the value of λ (and therefore none to the underlying probability distribution over z). First note that such a change applied to the scale of V doesn't change how weighted the associated ambiguity is to positive y values. It doesn't change "how far" $V(x^1) - V(x^2)$ is from zero, on average. This "weight to positive y values" is reflected in the value of $|\Lambda_f|$ (which is invariant with respect to such rescalings), and therefore (by Thm. 7(ii)) is also reflected in the value of

[40] Low learnability is not only a problem for agents with poor learning algorithms. Even for a Bayes-optimal learning algorithm, if the "signal to noise" of the private utility is poor, then the agent's intelligence for the actual r at hand can readily be far less than 1. (Bayes-optimality only means that x is set to maximize $E(\underline{g}_s \mid n, x)$, not to maximize $\underline{g}_s(x, r)$.)

[41] Trivially, the condition in Thm. 6 holds for (l', V_{a^*}, l, V_b) if it does for (l', V_a, l, V_b). In addition, $\Lambda_f(V_{a^*}; l', x^1, x^2) = \Lambda_f(V_a; l', x^1, x^2)$ while $K_{l', V_{a^*}, l, V_b} = \beta K_{l', V_a, l, V_b}$.

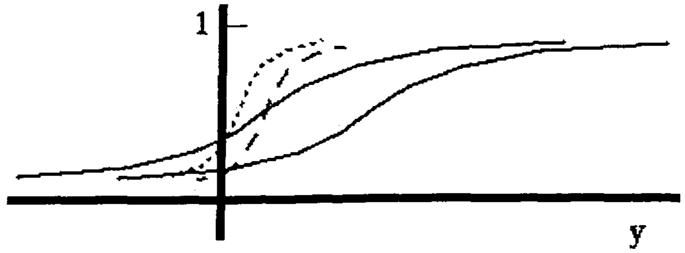

Figure 2: The leftmost solid line shows an ambiguity $A(y; V; l, x^1, x^2)$. The dotted line shows $A(y; V'; l, x^1, x^2)$ for $V' = aV$, $0 < a < 1$. $K_{V',V} = a$, and learnability of V' is the same as V's. The dashed line shows the dotted line right-shifted by $t_U(x^1, x^2)[h(x^1) - h(x^2)] > 0$, i.e., the ambiguity $A(y; U; l, x^1, x^2)$ for $U \equiv aV + h$. (Since we have not changed s, Thm. 6 must apply.) $\Lambda_f(U; l, x^1, x^2) > \Lambda_f(V'; l, x^1, x^2)$. Finally, the rightmost solid line depicts the dotted line expanded back to the scale of the leftmost solid line, i.e., the ambiguity of $U' \equiv \beta U$ where $\beta = 1/K_{V',V}$, so that $K_{U',V} = 1$. As with the previous one, this rescaling from W to T does not affect the learnability.

$t_U(x^1, x^2)\left[\frac{h(x^2)-h(x^1)}{K}\right]$. However such a rescaling can still be useful in how it "stretches" the CDF. To see how, note by Thm. 8(iii) that if V has better learnability than some other utility U, such stretching of V may provide a new utility V' such that in addition $K_{V',U} = 1$, which means that V' has better ambiguity than U (in light of Thm. 8(iii)).[42] In other words, to change the learnability we must induce a rightward offset in the (potentially scaled) ambiguity of V. Having done that, a subsequent rescaling can give us an aggregate K equal to 1 (without changing learnability), and thereby provide a final utility whose ambiguity lies everywhere below that of U. The value of that offset is given by the (β-independent) ambiguity shift. (See Fig. 2.)

(ii) Learnability and term 3

Plug Thm. 7 into Thm. 6, with U in Thm. 6 set to the x-ordering given by $V_b(., r)$. This shows that after appropriate rescaling of V_a, the triple (V_a, l') has better ambiguity than does (V_b, l) if it has better learnability.[43] If we plug that fact into Coroll. 1, we establish the following:

Corollary 4 *Fix r, l, l', V_a and V_b, where $\lambda \subseteq \nu$, as usual. Say $\exists K \in \Re^+, h : \xi \to \Re$, such that $\forall$ x^1, x^2*

i) $P_{V_a}(y^1, y^2; l', x^1, x^2) = P_{KV_b+h}(y^1, y^2; l, x^1, x^2)$;

and

[42] Note that such rescaling amounts to changing the temperature parameter in a Boltzmann learning algorithm.

[43] Note that this rescaling is done before we invoke the third premise. In this way we will be able to exploit that premise to do rescaling without invoking the assumption in Thm. 8(iii).

ii) $t_{V_b(.,r)}(x^1, x^2)\Lambda_f(V_a; l', x^1, x^2) > t_{V_b(.,r)}(x^1, x^2)\Lambda_f(V_b; l, x^1, x^2)$.

Then by appropriately rescaling V_a we can assure that

$$E^{[V_a;\lambda]}(N_{\rho,V_b} \mid r, l') \geq E^{[V_b;\lambda]}(N_{\rho,V_b} \mid r, l).$$

Consider changing the private utility from V_b to a V_a which is factored with respect to V_b. Then Coroll. 4 means that if this increases the learnability (in the x-ordering preferred by $V_b(., r)$) of one's private utility, then typically it results in higher expected intelligence, for the optimal scaling of that private utility. More precisely, express Coroll. 4 for $\lambda = \nu \cap \sigma$ and $l = l' \equiv (n, s)$ and then plug it into Coroll. 3 with $s^b = s, \underline{g}_{s^a} = V_a$ and $\underline{g}_{s^b} = V_b$, where s^a and s^b differ only in the associated private utility for our agent, and V_a and V_b are mutually factored. Then we see that if learnability is higher with s^a than with s^b (in the x-ordering preferred by $V_b(., r)$) for enough of the n for which $P(n \mid r, s^b)$ is non-negligible, then $s = s^a$ gives a higher expected intelligence conditioned on r and s than does $s = s^b$ (each intelligence evaluated for the associated optimal scale of the private utility).

As an added bonus, often the higher the learnability of a private utility, the more "slack" there is in setting the parameters of the associated learning algorithm while still having an ambiguity that's below that of some benchmark, low-learnability private utility. In other words, the higher the learnability, the less careful one must be in setting such parameters in order to achieve expected intelligence above some threshold. In particular, the greater the ambiguity shift in Coroll. 4, the broader the range of scales β for which βV_a has greater expected intelligence than does V_b. So by using private utilities with increased learnability often it becomes less crucial that one exactly optimize the learning algorithm's internal parameter setting the scale at which the algorithm examines utility values. This phenomenon can be amplified via "construction interference", for example as in the following result.

Corollary 5 *Fix r and two sets of utility-(λ-value) pairs, $\{V_t, l_t\}$ and $\{V^*, l_{t^*}\}$, indexed by t and t^*, respectively. Assume all quintuples $(r, l_{t^*}, V^*, l_t, V_t)$ obey Coroll. 4(i),(ii) with $V_a = V^*, V_b = V_t$, etc. For pedagogical simplicity, also take*

$\mathrm{sgn}[V_b(x^1, r) - V_b(x^2, r)] = \mathrm{sgn}[V_a(x^1, r) - V_a(x^2, r)] \equiv m,$

$\mathrm{sgn}[E(V_b^1 - V_b^2; l, x^1, x^2)] = \mathrm{sgn}[E(V_a^1 - V_a^2; l', x^1, x^2)] \equiv m',$

and $m = m'$.

i) Define

$$\begin{aligned}
\Delta_{t,t^* x^1,x^2} &\equiv \{\Lambda_f(V^*; l_{t^*}, x^1, x^2) - \Lambda_f(V_t; l_t, x^1, x^2)\}\sqrt{\int \mathrm{d}x f(x) Var(V_t; l_t, x)}, \\
B_{t,x^1,x^2} &\equiv \min(y : A(y; V_t(., r), V_t; l_t, x^1, x^2) = 1), \\
D_{t,x^1,x^2} &\equiv \max(y : A(y; V_t(., r), V_t; l_t, x^1, x^2) = 0),
\end{aligned}$$

where as usual f is a fixed but arbitrary distribution over x, and we assume $\Delta_{t,t^*,x^1,x^2} \geq 0 \; \forall t, t^*, x^1, x^2$.

ii) Define $K_{t,t^*} \equiv K_{l_{t^*},V^*,l_t,V_t}$*, and then define the subintervals of* $\Re$ *(one for each* (t, x^1, x^2) *triple)*

$$\begin{aligned}
L_{t,t^*,V^*,x^1,x^2} &\equiv \frac{1}{K_{t,t^*}}[\frac{B_{t,x^1,x^2}}{B_{t,x^1,x^2} + \Delta_{t,t^*,x^1,x^2}}, \frac{D_{t,x^1,x^2}}{D_{t,x^1,x^2} + \Delta_{t,t^*,x^1,x^2}}] \\
&\qquad \textit{if}\, D_{t,x^1,x^2} < -\Delta_{t,t^*,x^1,x^2}, \\
&\equiv \frac{1}{K_{t,t^*}}[\frac{B_{t,x^1,x^2}}{B_{t,x^1,x^2} + \Delta_{t,t^*,x^1,x^2}}, \infty) \\
&\qquad \textit{if} - \Delta_{t,t^*,x^1,x^2} \leq D_{t,x^1,x^2} < 0, \\
&\equiv \frac{1}{K_{t,t^*}}[\frac{D_{t,x^1,x^2}}{D_{t,x^1,x^2} + \Delta_{t,t^*,x^1,x^2}}, \infty) \\
&\qquad \textit{otherwise},
\end{aligned}$$

and

$$L_{t,t^*,V^*} \equiv \cap_{x^1,x^2} L_{t,t^*,V^*,x^1,x^2}.$$

iii) Define $L_{t^*,V^*} \equiv \cup_t L_{t,t^*,V^*}$.

Then for every t^*, $\forall \beta \in L_{t^*,V^*}$,

$$E^{[\beta V^*;\lambda]}(N_{V^*} \mid r, l_{t^*}) \geq min_t E^{[V_t;\lambda]}(N_{V_t} \mid r, l_t).$$

Note that $B_{t,x^1,x^2} \geq 0$ always, since $m = m'$ for (l_t, V_t). Accordingly, L_{t^*,V^*} is never empty, always containing $\cup_t \frac{1}{K_{t,t^*}}$ at least.[44,45]

To help put Coroll. 5 in context, apply Coroll. 4 to the scenario of Coroll. 5. This establishes that for any t^*, $\exists \beta \in L_{t^*,V^*}$ such that $E^{[\beta V^*;\lambda]}(N_{\beta V^*} \mid r, l_{t^*}) \geq max_t E^{[V_t;\lambda]}(N_{\beta V_t} \mid r, l_t)$. Note also the immediate implication of Coroll. 5 that $\forall \beta \in \cap_{t^*} L_{t^*,V^*}$,

$$min_{t^*} E^{[\beta V^*;\lambda]}(N_{V^*} \mid r, l_{t^*}) \geq min_t E^{[V_t;\lambda]}(N_{V_t} \mid r, l_t).$$

As an example of Coroll. 5, take $\lambda = \nu \cap \sigma$, have l_{t^*} equal some fixed l^* $\forall t^*, V^* \equiv \underline{g}_{s^*}$, and $V_t \equiv \underline{g}_{s_t}$ $\forall t$. Have real-valued $t \in [t_1 > 0, t_2]$, where $V_t = V_{t_1} \frac{t}{t_1}$. So assuming $\Lambda_f(V^*; l^*, x^1, x^2) \geq \Lambda_f(V_t; l_t, x^1, x^2)$ $\forall x^1, x^2$ as usual, the range in the logarithms of β for which $E^{[\beta V^*;\lambda]}(N_{V^*} \mid r, l^*) \geq min_t E^{[V_t;\lambda]}(N_{V_t} \mid r, l_t)$ is greater than or equal to $\ln(t_2)$ - $\ln(t_1)$.[46]

[44] If (unlike in Coroll. 4) the value of K can change with the (x^1, x^2) values, then those indices must be added to K's subscripts. In this case the conclusion of Coroll. 4 need not hold; L_{t^*,V^*} can be empty.

[45] A subtle point is that in situations where $D_{t,x^1,x^2} > 0$, we can increase the scale of V_t as many times as we want and assuredly improve its ambiguity each time. (This is not something we can do in the other situations.) Accordingly, if *every* instance going into L_{t^*,V^*} is such a situation, then our conclusion that rescaling V^* can assuredly give better expected intelligence than V_t is a bit irrelevant; in this scenario we can also rescale V_t to assuredly improve *its* expected intelligence.

[46] To see this, note that t sets the scale of V_t, just like β does for V^*. Furthermore, $K_t \equiv K_{t,t^*} = \frac{t_1 K_{t_1,t^*}}{t}$ if $P(r'; l^*) = P(r'; l_t)$ $\forall r', t$ (cf. Thm. 8(iii)). So $1/K_t$, which we know is contained in L_{t,t^*,V^*}, equals $\frac{t}{t_1 K_{t_1}}$. Now apply Coroll. 5.

As another example, choose $\{l_t\} = \{l_{t^*}\} = \{n \in \text{supp}\, P(\nu \mid r, s^i), s^i_\rho)\}$ for some set of σ values $\{s^i_\rho\}$, with $V_t = V_{n,s^i} = \underline{g}_{s^i}\ \forall i$. Also presume that $\forall \beta$, there is a design coordinate value s^β_ρ such that $\underline{g}_{s^\beta} = \beta V^*$. If we now plug the conclusions of Coroll. 5 into Coroll. 3, we establish that $\forall i, \beta \in \cap_{n \in \text{supp}\, P(\nu|r,s^i)} L_{n,s^i_\rho,V^*}$,

$$E(N_{\underline{g}_{s^\beta}} \mid r, s^\beta_\rho) \geq E(N_{\underline{g}_{s'}} \mid r, s^i),$$

and therefore $\forall \beta \in \cap_{s^i, n \in \text{supp}\, P(\nu|r,s^i)} L_{n,s^i_\rho,V^*}$,

$$E(N_{\underline{g}_{s^\beta}} \mid r, s^\beta_\rho) \geq min_{s^i, n \in \text{supp}\, P(\nu|r)} E(N_{\underline{g}_{s^i}} \mid r, s^i)$$

(iii) Aristocrat Utility

In general, there is no utility that is both factored with respect to the world utility and has infinite learnability.[47] The following result allows us to solve for the private utility that maximizes learnability, and thereby find the private utility for agent ρ that should give best performance under the first three premises:

Theorem 9

i) A utility U_1 is factored with respect to U_2 at z iff $\forall\ z' \in \rho(z) \equiv r$, with $x \equiv \xi(z')$, $U_1(x,r) = F_r(U_2(z')) - D(r)$, for some function D and some r-parameterized function F_r with positive derivative.

ii) For fixed $l \in \lambda \subseteq \nu$, r, x^1, x^2, and F, the D that maximizes $\Lambda_f(U_1; l, x^1, x^2)$ is the (l, x^1, x^2)-independent quantity $E_{f(x)}(F_r(U_2(\xi, r)))$.

iii) The f that maximizes the associated ambiguity shift between U_2 and U_1 is

$$\text{argmin}_f \left[\frac{E_{f(x)}\{\text{Var}(U_2; l, \xi)\}}{E_{f(x^1),f(x^2)}\{\text{Var}((F^1 - F^2)\delta(r^1 - r^2); l, \xi^1, \xi^2)\}} \right],$$

where the subscript on the denominator expectation indicates that both x's are averaged according to f, and the delta function there means that our two F's (one for each x) are evaluated at the same r.

A particularly important example of a function F_r meeting the condition in Thm. 9 is $F_r(U_2) = U_2$. This choice results in the difference utility U_1 that takes $z = (x, r) \rightarrow U_2(x, r) - E_f(U_2(\xi, r))$. We call this the **Aristocrat Utility** (AU)

[47] As an example of when having both conditions is impossible, take $r \in \{r^1, r^2\}$, $x \in \{x^1, x^2\}$, and $G(x^1, r^1) > G(x^2, r^1)$, while $G(x^2, r^2) > G(x^1, r^2)$. Then by Thm. 1, we also must have $\underline{\gamma}_\rho(x^1, r^1) > \underline{\gamma}_\rho(x^2, r^1)$ and $\underline{\gamma}_\rho(x^2, r^2) > \underline{\gamma}_\rho(x^1, r^2)$. Also assume that $P(r'; l) = \delta(r' - r)\ \ \forall\, r, s$, so $P(U = u; l, x) = \delta(u - U(x, r))$ always.

Define $A \equiv \underline{\gamma}_\rho(x^1, r^2) - \underline{\gamma}_\rho(x^1, r^1)$, $C \equiv \underline{\gamma}_\rho(x^2, r^2) - \underline{\gamma}_\rho(x^1, r^2)$, $B \equiv \underline{\gamma}_\rho(x^2, r^1) - \underline{\gamma}_\rho(x^2, r^2)$, and $D \equiv \underline{\gamma}_\rho(x^1, r^1) - \underline{\gamma}_\rho(x^2, r^1)$. So $A + B + C + D = 0$, and both $C > 0$ and $D > 0$.

Take $f(x) = 1/2$ for both x, so $\int dx\, f(x)\, \text{Var}(U; r, s, x) = [A^2 + B^2]/4$, which by convexity $\geq [(A + B)/2)]^2 = [(C + D)/2)]^2$. In turn, $[E(U; l, x^1) - E(U; l, x^2)]^2 = [(D - C)/2]^2 \leq [(C + D)/2)]^2$. Combining, by the definition of learnability we see that it is bounded above by 1. **QED**.

for U_2 at z, $\mathrm{AU}_{U_2,f}(z)$, reflecting the fact that it is the difference between the value of U_2 at the actual z and the average such utility.

Say a particular choice of f, f', results in conditions (i) and (ii) of Coroll. 4 being met with $V_b = U_2$ and $V_a = \mathrm{AU}_{U_2,f'}$, for the choice of λ etc. discussed just after the presentation of Coroll. 4. Then we know by that corollary that once it is appropriately rescaled, using the AU for U_2 as ρ's private utility results in an expected intelligence with that is larger than is the expected intelligence that arises from using U_2 as the private utility. (Note that U_2 and $\mathrm{AU}_{U_2,f'}$ are mutually factored.) Moreover, by Thm. 9 any other difference utility that obeys Coroll. 4(i)(ii) (in concert with U_2) must have worse ambiguity than does $AU_{U_2,f'}$, and therefore worse expected intelligence.[48]

To evaluate AU for some G at some z we must be able to list all $z' \in \rho(z)$. This can be a major difficulty, for example if one cannot observe all degrees of freedom of the system. Even if we can list all such z', we must also be able to calculate G for all those z', an often daunting task which simple observation of the actual $G(z)$ at hand cannot fulfill (in contrast to the calculational needed with a team game, for example).

Even when we cannot calculate an AU exactly though, we can often use an approximate AU and thereby improve performance over a team game. For example, in an iterated game, at timestep t, r for a particular player i reflects the state of the other players it is confronting. In such a situation, by observing r, often we can approximate $E_f(\underline{g}_i(\xi, r))$ by an appropriate average of the value of $\underline{g}_i$ over those preceding iterations when the state of the other players was r, with f being the frequency distribution of moves made by i in those iterations. In particular, consider a "bake-off" tournament of a 2-player game in which each player in the tournament plays one other player in each round, and keeps track of who it has played in the past and with what move and resultant outcome. In such a situation, the expectation value for player i confronting player j that gives $\mathrm{AU}_{\underline{g}}$ can often be approximated by the average payoff of player i over those previous runs where i's opponent was j.

On the other hand, even when we can evaluate AU exactly, it may be that the conditions in Coroll. 4 are badly violated. In such situations increasing learnability by using AU will not necessarily improve expected intelligence, and accordingly AU may not induce optimal performance. Indeed, it may induce *worse* performance than the team game in such situations. On the other hand, there are other modifications to the private utility that (under the first premise) may improve expected intelligence in these situations. An example of such a utility is the CU, as illustrated in [22].

(iv) Wonderful Life Utility

One technique that will often circumvent the difficulties in evaluating AU is to replace ρ with a coarser partition, having poorer resolution. While this replace-

[48] Note though that in general there may be a utility $F_r(U_2) - D(r)$ with better learnability than AU, for example if F_r is non-linear. Note also that whether $AU_{U_2,f'}$ obeys conditions 4(i)(ii) will depend on the choice of f', in general.

ment usually decreases learnability below that of AU, it still results in utilities that are far more learnable than team game utilities, while (like team games) not requiring knowledge of the set of worldpoints $\rho(z)$ in full. In this subsection we illustrate making such a replacement for difference utilities.

We concentrate on the case where the domain of the lead utility D_1 is all of ζ, and the secondary utility $D_2 = D_1(\phi(z))$ for some function $\phi : \zeta \rightarrow \zeta$ where $\forall z \in C, \phi$ depends only on r, i.e., $\forall\, r$, $\forall z'$, $z'' \in r$, $\phi(z') = \phi(z'')$. So specifying the utility consists of choosing ϕ. While in general we can make the choice that best suits our purposes, here we will only consider a particular class of ϕ's. A more general approach might, for example, choose ϕ to maximize learnability. Intuitively, the resulting difference utility is equivalent to subtracting D_1 of a transformed z from the original $D_1(z)$, with the transform chosen to maximize the signal-to-noise of the resultant function. See the discussion of Thm. 7.

Let π be a partition of ζ. Fix some subset of ζ called the **clamping element** $\mathrm{CL}_{\vec{\pi}}$ such that $\forall p \in \pi, D_1$ is invariant across the (assumed non-empty) intersection of $\mathrm{CL}_{\vec{\pi}}$ and p.[49] Define an associated projection operator $\mathrm{CL}_{\vec{\pi}}(z) \equiv \mathrm{CL}_{\vec{\pi}} \cap \pi(z)$, which for any $p \in \pi$ maps all worldpoints lying in p to the same subregion of that element, a subregion having a constant D_1 value.[50] Then the **Wonderful Life Utility** (WLU) of D_1 and π is defined by

$$\mathrm{WLU}_{D_1,\pi}(z) \equiv D_1(z) - D_1(\mathrm{CL}_{\vec{\pi}}(z)).^{51}$$

To state our main theorem concerning WLU, for any partition of ζ, π, and any set $B \subseteq \zeta$, define $B \cap \pi$ to be a partition of B with elements given by the intersections of B with the elements of π. Furthermore, recall from App. B that given two partitions π_1 and π_2, $\pi_1 \subseteq \pi_2$ iff each element of π_1 is a subset of an element of π_2. Then the following holds regardless of what subset of ζ forms C:

Theorem 10 *Let π and $\pi' \subseteq \pi$ be two partitions of ζ. Then $\mathrm{WLU}_{D_1,\pi}$ is factored with respect to D_1 for coordinate $C \cap \pi'$ $\forall\, z \in C$.*

As an example, with $\rho \equiv C \cap \pi$, $\mathrm{WLU}_{G,\pi}$ is factored with respect to G for coordinate ρ.

Note that $\pi' \subseteq \pi$ means that π' is either identical to π or a "finer-resolution" version of π. So $z \rightarrow \mathrm{CL}_{\vec{\pi}} \cap \pi(z)$, by sending all points in $\pi(z)$ to the same point, is a more severe operation, resulting in a greater loss of information, than is $z \rightarrow \mathrm{CL}_{\vec{\pi} \cap \pi'(z)}$, which can map different points on $\pi(z)$ differently. So Thm. 10 means we can err on the side of being over-severe in our choice of clamping operator and the associated WLU is still factored.[52]

[49] Note that $\mathrm{CL}_{\vec{\pi}}$ automatically has this property, independent of D_1, if its intersection with each element of π consists of a single worldpoint.

[50] Note that both $\mathrm{CL}_{\vec{\pi}}$ and $\mathrm{CL}_{\vec{\pi}}(z)$ are implicitly parameterized by D_1.

[51] Note that if there is some $x' \in \xi$ such that $\mathrm{CL}_{\vec{\pi}}(x, r) = (x', r)$ $\forall\, x, r$, then WLU is a special type of AU, with a delta function f.

[52] Sometimes $\mathrm{WLU}_{G,\pi'}(z)$ will be factored with respect to G for coordinate $C \cap \pi$ even though $\pi' \subseteq \pi$. For example, this is the case if G is independent of precisely which of the elements of π' contains z, so long as all of those elements are in $\pi(z)$. However in general

There are other advantages to WLU that hold even when $\pi = \pi'$. For example, in general $\mathrm{CL}_{\hat{\pi}}(z)$ need not lie on the set C (n.b., π and $\hat{\pi}$ are partitions of ζ, not C). In such a case the function $G(\mathrm{CL}_{\hat{\pi}}(z)) : C \rightarrow \Re$ is not specified by the function $G(z) : C \rightarrow \Re$. In this situation we are free to choose the values $G(\mathrm{CL}_{\hat{\pi}}(z))$ to best suit our purposes, e.g., to maximize learnability.

An associated advantage is that to evaluate the WLU for coordinate $C \cap \pi$, we do *not* need to know the detailed structure of C. This is what using WLU for the coarser partition π rather than the AU for the original coordinate $C \cap \pi'$ gains us. Given a choice of clamping element, so long as we know $G(z)$ and $\pi(z)$, together with the functional form of G for the appropriate subsets of ζ, we know the value of $\mathrm{WLU}_{G,\pi}(z)$. These advantages are borne out by the experiments reported in [17].

(v) WLU in repeated games

As an example of WLU, say we have a deterministic and temporally invertible repeated game (see App. D). Let the $\{\omega_1, \omega_2, \ldots, \omega_J\}$ and $\{\theta_1, \theta_2, \ldots, \theta_L\}$ be two sets of generalized coordinates of C^T (not necessarily repeating coordinates). Consider a particular player/agent, and presume that $\forall t'$ there is a single-valued mapping from $r^{t'} \rightarrow (w_1, w_2, \ldots, w_J)$, and one from $(x^{t'}, r^{t'}) \rightarrow (q_1, q_2, \ldots, q_L)$ (both implicitly set by C). So the player's context at time t' fixes the values of the ω_i (defined for time T), and by adding in the player's move at that time we also fix the values of the θ_i. Say we also have a utility U that is a single-valued function of $(w_1, w_2, \ldots, w_J, q_1, q_2, \ldots, q_L)$.

Take π to be the partition of ζ whose elements are specified by the joint values of the $\{\omega_1, \omega_2, \ldots, \omega_J\}$. Take $\mathrm{CL}_{\hat{\pi}}$ to be a set of z sharing some fixed values of $\{\theta_1, \theta_2, \ldots, \theta_L\}$. Note that U is constant across the intersection of $\mathrm{CL}_{\hat{\pi}}$ with any single element of π, as required for it to define a WLU.

Intuitively, $\mathrm{CL}_{\hat{\pi}}(z)$ is formed by "clamping" the values of the $\{\theta_1, \theta_2, \ldots, \theta_L\}$ to their fixed value while leaving the $\{\omega_1, \omega_2, \ldots, \omega_J\}$ values unchanged. Moreover, since $r^{t'} \rightarrow (w_1, w_2, \ldots, w_J)$ is single-valued, we know that any dependency of the important aspects of z (as far as U is concerned) on our player's move at time t' is given by (a subset of) the values $\{q_1, q_2, \ldots, q_L\}$. (Recall that all values $x^{t'}$ are allowed to accompany a particular $r^{t'}$.)

Now by Thm. 10, we know that $\mathrm{WLU}_{U,\pi}$ is factored with respect to U for coordinate $C \cap \pi'$ for any partition π' that is a refined version of π. In addition, $\rho^{t'} \subseteq \pi$. So $\mathrm{WLU}_{U,\pi}$ is factored with respect to U for the coordinate given by $C \cap \rho^{t'} = \rho^{t'}$, i.e., it is factored for our player's context coordinate at time t'.

When the $\{\theta_i\}$ are minimal in that none of them is a single-valued mapping of $r^{t'}$ (i.e., none can be transferred into the set of $\{w_i\}$), we say they are our

such factoredness will not hold. Even if it doesn't though, say G is *relatively* insensitive to which of the elements of π' contains z, over the set of all such elements that are in $\pi(z)$. Then $\mathrm{WLU}_{G,\pi'}(z)$ will be quite close to factored for coordinate $C \cap \pi$. This often allows us to be "sloppy" in using WLU's, by taking π' to be only those degrees of freedom $C \cap \pi$ with "significant impact" on the value of G.

player's **effect set** [17].[53] Often a player's behavior can be modified to ensure that a particular set of $\{\theta_i\}$ contains its effect set for some particular time. When we can do this it will assure that some associated variables $\{\omega_i\}$ specify (a partition π that gives) a $\mathrm{WLU}_{G,\pi}$ for our player's move at that time that is factored with respect to G.

(vi) WLU in large systems

Consider the case of very large systems, in which G typically depends significantly on many more degrees of freedom than can be varied within any single element of ρ (i.e., depends more on the value of r than on where the system is within that r). So we can write $G(x,r) = G_1(x,r) + G_2(r)$ where the values of G_2 in C are far greater than those of G_1, and correspondingly the changes in the value of G_1 as one moves across C are far smaller than those of G_2. In such cases, with $\rho = C \cap \pi$ as usual, the learnability of G is far less than that of $\mathrm{WLU}_{G,\pi}$. This is due to the following slightly more general theorem:

Theorem 11 *Let κ and $\pi \subseteq \kappa$ be two partitions of ζ. Write $H(z) = H_1(z) + H_2(\kappa(z))$, where H is defined over all ζ, and consider the agent $\rho = C \cap \pi$. Fix $l \in \lambda \subseteq \nu$, and define*

$$M \equiv \{\max_{z,z'}[H_1(z) - H_1(z')]\}^2$$

and

$$L \equiv \int \mathrm{d}k'\,\mathrm{d}k''\,P(k';l)P(k'';l)[H_2(k') - H_2(k'')]^2.$$

Then independent of f, $\mathrm{CL}\text{-}_{\kappa}$, x^1 and x^2,

$$\frac{\Lambda_f(\mathrm{WLU}_{H,\kappa};l,x^1,x^2)}{\Lambda_f(H;l,x^1,x^2)} \geq \frac{L}{2M} - \sqrt{\frac{L}{M}}.$$

Note that as κ becomes progressively coarser and coarser, L shrinks. So such coarsening of the clamping element will typically lead to worse learnability. In fact, in the limit of $\kappa = \emptyset$, $\mathrm{WLU}_{H,\kappa}$ just equals H minus a constant. So in that

[53] Sometimes the $(q_1, q_2, \ldots, q_n)$ value specifying the clamping element of an effect set can intuitively be viewed as a "null action", so that clamping can be viewed as "removing agent ρ from the system". Intuitively, in this case we can view WLU as a first order subtraction from G of the effects on it of specifying those degrees of freedom *not* contained in the effect set (hence the name "wonderful life" utility—c.f. the Frank Capra movie). More formally, in such circumstances WLU can be viewed as an extension of the Groves mechanism of traditional mechanism design, generalized to concern arbitrary (potentially time-extended) world utility functions, and to concern situations having nothing to do with valuation functions, (quasi-linear) preferences, types, revelations, or the like. (See [7, 2, 14, 2, 10, 16, 8, 27, 13].) Due to its concern for signal-to-noise issues though, this extension relies crucially on re-scaling of G. (Indeed, if one just subtracts the clamped term without any such re-scaling, ambiguity can be badly distorted, so that performance can degrade substantially [23].) In addition, this extension allows alternative choices of the clamping operator, even clamping to illegal (i.e., not $\in C$) worldpoints. This extension also can be used even in cases where there is no action that can be viewed as a "null action", equivalent to "removing the agent from the system".

limit, $\mathrm{WLU}_{H,\kappa}$ and H must have the exact same learnability — in agreement with Thm. 11 and the fact that $L = 0$ in that limit.

When L greatly exceeds M the bound in Thm. 11 is much greater than 1. So if we take $H = G$ and $\kappa = \pi$, Thm. 11 tells us that for very large systems, setting the private utility to G's WLU rather than to G may result in an extreme growth in learnability.[54] In particular, for $\lambda = \nu \cap \sigma$, in large systems it may be that $L >> M \forall\ l$ such that $P(l \mid s)$ is non-infinitesimal. Under the first three premises, assuming $\mathrm{WLU}_{G,\kappa}$ and G obey the conditions in Coroll. 4(i),(ii), this means that setting the private utility to WLU will result in larger expected intelligence of the agent than will setting it to G. Moreover, since that WLU is factored with respect to G, this improvement in term 3 of the central equation will not be accompanied by a degradation in term 2. This ability to scale well to large systems is one of the major advantages of WLU and AU.

(vii) WLU in spin glasses

As a final example, consider a spin glass with spins $\{b_i\}$. For each spin i let $\vec{b}_{-i}$ be the set of spins other than i, and for each i let h_i and F_i be any two functions such that the Hamiltonian can be written as $\mathcal{H}(\vec{b}) = h_i(b_i, \vec{b}_{-i}) + F_i(\vec{b}_{-i})$. In particular, for $\mathcal{H}(\vec{b}) = \sum_{jk} \mathcal{H}_{jk} b_j b_k + \sum_j \mathcal{H}_j b_j$, we can have $F_i(\vec{b}_{-i}) = \sum_{j \neq i, k \neq i} \mathcal{H}_{jk} b_j b_k + \sum_{j \neq i} \mathcal{H}_j b_j$, and $h_i(b_i, \vec{b}_{-i}) = \mathcal{H}_i b_i + \mathcal{H}_{ii} b_i^2 + \sum_{j \neq i} [\mathcal{H}_{ij} + \mathcal{H}_{ji}] b_j b_i / 2$. Since at equilibrium $\vec{b}$ minimizes $\mathcal{H}$, and therefore given the equilibrium value of $\vec{b}_{-i}$, at the $\mathcal{H}$-minimizing point b_i is set to the value that minimizes $h_i(b_i, \vec{b}_{-i})$.

We can view this as an instance of a collective where $\mathcal{H}$ is the (negative) world utility G for a system of "agents" ρ with move b_ρ, and $g_\rho = h_\rho$. For all ρ, at the $\vec{b}$ that maximizes G, b_ρ is set to the value that maximizes $-h_\rho$ given $\vec{b}_{-\rho}$. More generally, $h_i(b_i, \vec{b}_{-i}) = \mathcal{H}(\vec{b}) - F_i(\vec{b}_{-i})$ is factored with respect to $G(\vec{b})$ (cf. Thm. 2), with the context for each agent ρ being $\vec{b}_{-\rho}$ and $C = \zeta$ being the set of all vectors $\vec{b}$. So any $\vec{b}$ (locally) maximizing G also simultaneously maximizes all of the $-h_i$. Frustration then is a state where all the agents' intelligences equal 1, but the system is at a local rather than global maximum of G.

Consider a particular spin/agent, ρ. Embed C, the set of all possible $\vec{b}$, in some larger space that allows the spin ρ to take on additional values, and redefine ζ to be that larger space. Let π be an associated ζ-partition such that $\rho \equiv C \cap \pi$. Take $\mathrm{CL}_{\cdot\pi}$ to be some set off of C. Extend the domain of definition of h_ρ by setting $h_\rho(\mathrm{CL}_{\cdot\pi}(\vec{b}) = 0\ \ \forall\, \vec{b}_\rho \notin C$. Then $\mathrm{WLU}_{G,\pi} = -h_\rho$, i.e., WLU is the "local Hamiltonian" perceived by spin ρ, whereas G is the Hamiltonian of the entire system.

So by Thm. 11, if the number of nonzero coupling strengths between ρ and the other spins is much smaller than the total number of nonzero coupling strengths in the system, then the learnability of ρ's local Hamiltonian far exceeds

[54] Trivially, since learnability of AU is bounded below by that of WLU, its learnability must exceed that of a team game at least as much as WLU's does.

that of the global Hamiltonian. Accordingly, consider casting the evolution of the spin system as an iterated game, with each spin controlled by a learning algorithm, and each $\underline{g}_{\rho,s^t}$ set to either spin ρ's local Hamiltonian at time t, or to the global Hamiltonian at that time. (See App. D.) Then since WLU is factored with respect to G, we would expect (under the first three premises, and assuming conditions 4.1(i)(ii) hold, etc.) that at any particular timestep of the game $\vec{b}$ is closer to a local peak of the *global* Hamiltonian if the agents use the value at that timestep of their local Hamiltonians as their private utilities, rather than use the value of the global Hamiltonian at that timestep.

If we also incorporate techniques addressing term 1 in the central equation, then we can ensure that such local peaks are large compared to the global peak. Moreover, if we have the spins use a WLU with better learnability, we would expect faster convergence still. Similarly, if the spins use AU rather than their local Hamiltonians, then since this increases learnability, performance of the overall system should improve further still. (Roughly speaking, such a change in private utilities is equivalent to having the agents use mean-field approximations of their local Hamiltonians as their rewards rather than the actual values of their local Hamiltonians.) More generally, any modification of the system that induces higher learnability (while maintaining factoredness of the individual spins' private utilities with respect to the original Hamiltonian) should result in faster convergence to the minimum of the original Hamiltonian. The foregoing is borne out in experiments reported in [24].

Acknowledgements

I would like to thank Mike New, John Lawson, Joe Sill, Peter Stone, and especially Kagan Tumer and Mark Millonas for helpful discussion.

A Intelligence, Percentiles and Generalized CDF's

A useful example of intelligence is the following:

$$N_{\rho,U}(z) \equiv \int \mathrm{d}\mu_{\rho(z)}\,\Theta[U(z) - U(z')] \qquad \text{(A.1)}$$

with the subscript on the (usually normalized) measure indicating it is restricted to $z' \in \rho(z)$ (usually it is also nowhere-zero in that region). For consistency with its use in expansions of CDF's, the Heaviside function is here taken to equal 0/1 depending on whether its argument is less than 0 or not. (Having $\Theta(0) = 0$ in Eq. A.1 is also a valid intelligence operator.) Intuitively, this kind of intelligence quantifies the performance of z in terms of its percentile rank, exactly as is conventionally done in tests of human cognitive performance. Note that this type of intelligence is a model-free quantification of performance quality; even if z is set by an agent that wants large $N_{\rho,U}$ and $N_{\rho,U}(z)$ turns out to be large "by luck", we still give that agent credit. The analogous coordinateless expression is given by $N_U(z) \equiv \int \mathrm{d}\mu(z')\,\Theta[U(z) - U(z')]$ where μ runs over all of C.

There is a close relationship between CDF's and intelligence in general, not just percentile-based intelligence. Thm. 3 provides an example of that relationship. For percentile-based intelligence though the relationship is even deeper. In particular, coordinateless percentile-based intelligence can be viewed as a generalization of cumulative distribution functions (CDF's). This generalization applies to arbitrary spaces serving as the argument of the underlying probability density function (not just $\mathfrak{R}^1$) and does not arbitrarily restrict the "sweep direction" (said direction being from $-\infty$ to $+\infty$ for the conventional case). In particular, for the special case of $z \in \mathfrak{R}^n$ and invertible $U(.)$ where $|\nabla_z U(z)| = 1$ a.e., $|\nabla_z N_U(z)|$ gives the probability density $\mu(z)$ and $0 \leq N_U(z) \leq 1 \;\; \forall \; z$, just like with the conventional CDF for which the underlying space is $\mathfrak{R}^1$. (In fact, for $U(z \in \mathfrak{R}^1) = z +$ constant, $N_U(z)$ is identical to the conventional CDF of the underlying distribution $\mu(z)$.) For the more general case, intuitively, U itself provides the flow lines of the sweep.

Percentile-type intelligence is arbitrary up to the choice of measure μ, and in a certain sense essentially any intelligence (in the sense defined in the text) can be "expressed" as a percentile-type intelligence. As an alternative to these kinds of intelligences, one might consider standardizing a utility U by simply subtracting some canonical value (like the expected value of U) from $U(z)$. This operation doesn't take into account the width of the distribution over U values however, and therefore doesn't tell us how significant a particular value $U(z) - E(U)$ is. To circumvent this difficulty one might "recalibrate" $U(z) - E(U)$ by dividing it by the variance of the distribution, but this can be misleading for skewed distributions; higher-order moments may be important. Formally, even such a recalibrated functions runs afoul of condition (i) in the definition of intelligence.

One important property of percentile-type intelligence is that with uncountable ζ and a utility U having no plateaus in ζ, if $P(\hat{}r \mid r, s) = \mu_r(\hat{}r)$ and is independent of r, then $P(N_U(z) \mid s)$ is constant, regardless of U and μ. More formally,

Theorem A.1 *Assume that for all y in some subinterval of $[0.0, 1.0]$, for all r in $\text{supp}\, P(. \mid s)$ there exists $\hat{}r$ such that the intelligence $N_{\rho,U}(r, \hat{}r) = y$. Restrict attention to cases where the intelligence measure $\mu_r(\hat{}r) = P(\hat{}r \mid r, s)$ and is independent of r. For all such cases, $P(N_U(z) \mid s)$ is flat with value 1.0, independent of both μ and U.*

Proof: We use the complement notation discussed in App. B. Write

$$P(N_{\rho,U}(r, \hat{}r) = y \mid s) = \int \mathrm{d}r\, \mathrm{d}\hat{}r'\, P(r \mid s) P(\hat{}r' \mid r, s) P(N_{\rho,U}(r, \hat{}r') = y \mid r, s)$$

Next write $P(N_{\rho,U}(r, \hat{}r) = y \mid r, s)$ as the derivative of the CDF $P(N_{\rho,U}(r, \hat{}r) \leq y \mid r, s)$ with respect to y. Now by assumption there exists a $\hat{}r$ such that $N_{\rho,U}(r, \hat{}r) = y$. So we can rewrite that CDF as

$$P(N_{\rho,U}(r, \hat{}r') \leq N_{\rho,U}(r, \hat{}r) \mid r, s),$$

where the probability is over $\hat{}r'$, according to the distribution $P(\hat{}r' \mid r, s)$.

We can rewrite this CDF as

$$P(U(r, \hat{}r') \leq U(r, \hat{}r) \mid r, s),$$

by property (ii) of the general definition of intelligence. In turn we can write this as

$$\begin{aligned}
&\int \mathrm{d}\hat{}r' \, P(\hat{}r'' \mid r, s)\Theta(U(r, \hat{}r) - U(r, \hat{}r')) \\
&\quad = \int \mathrm{d}\hat{}r' \, \mu(\hat{}r')\Theta(U(r, \hat{}r) - U(r, \hat{}r') \qquad \text{(by assumption)} \\
&\quad = N_{\rho, U}(r, \hat{}r) \qquad \text{(by definition of intelligence)} \\
&\quad = y.
\end{aligned}$$

Therefore the derivative of our CDF $= 1$. **QED**.

Intuitively, this theorem says that the probability that a randomly sampled point has a value of $\overline{U} \leq$ the y'th percentile of U is just y, so its derivative $= 1$, independent of the underlying distributions. Note that both the assumption that $P(\hat{}r \mid r, s)$ is independent of r and having $\mu(\hat{}r) = P(\hat{}r \mid s)$ is "natural" in single-stage games—but not necessarily in multi-stage games (see App. D).

If the conditions in the theorem apply, then choice of U is irrelevant to term 3 in the central equation. If we choose a "reasonable" U this means that we cannot have $P(\hat{}r \mid s) = \mu(\hat{}r)$ if we want to have choice of coordinate utility make a difference.

Note though that the assumption about the subinterval of $[0.0, 1.0]$ will be violated if U has isoclines of non-zero probability. This will occur if μ has delta functions, or if ζ is a Euclidean space and U has plateaus extending over the support of $P(z \mid s)$. A particular example of the former is when ζ is a countable space—the theorem does not apply to categorical spaces.

B Theory of Generalized Coordinates

It can be useful to view coordinates as "subscripts" on "vectors" z. Similarly, in light of their role as partitions of C, it can be useful to view separate coordinates as separate sets, complete with analogues of the conventional operations of set theory. As explicated in this appendix, these two perspectives are intimately related.

Now define $z_\rho \equiv \hat{}\rho(z)$, so $z_{\hat{}\rho} = \rho(z)$. Typically we identify the elements of z_ρ not by the sets making up $\hat{}\rho(z)$, but rather by the labels of those sets. This notation is convenient when ζ is a multi-dimensional vector space, since it makes the natural identification of contexts with vector components consistent with the conventional subscripting of vectors. For example, say $\zeta = \mathfrak{R}^3$, with elements written (x, y, z). Then a context for an "agent" making "move" x, ρ_x, is most naturally taken to be the partition of $\mathfrak{R}^3$ that is indexed by the moves of the other players, i.e., the values of y and z. In other words, specifying y and z gives a line delineating the remaining degrees of freedom of setting $z \in \Re^3$ that are available to agent x in determining its move, and each such line is an element of the partition ρ_x. For this ρ_x, we can take the complement $\hat{}\rho_x$ to be the partition of $\Re^3$ whose elements are planes of constant x, i.e., whose elements are labeled by the value of x. We can then write $\hat{}\rho_x(z) = z_{\rho_x} \equiv z_x$. With this choice z_x is just z's x value (recall we identify an element of z_x by its label). This is in accord with the usual notation for vector subscripts

To formulate a set theory over coordinates, first note that coordinates are not just sets, but special kinds of sets—a coordinate's elements are non-intersecting subsets of C whose union equals C. So for example to have $\rho_1 \cup \rho_2$ be a coordinate, it cannot be given by the set of all elements of ρ_1 and ρ_2, as it would under the conventional set theoretic definition of the union operator. (If the union operator were defined in that conventional manner, its elements would have non-zero intersection with one another.) This means that we cannot simply view coordinates as conventional sets and define the set theory operators over coordinates accordingly; we need new definitions.

To flesh out a full "set-theory" of coordinates, first note that the complement operation has already been defined. (Note that unlike in conventional set theory, here the complement operator is not single-valued.) We can also define the null set coordinate $\emptyset$ as the coordinate each of whose members is a single $z \in C$. So $\emptyset$ is bijectively related to ζ, and $\hat{}\emptyset$ can be taken to be the coordinate consisting of a single set: all of C.

To define the analogue of set inclusion, given two coordinates ρ_1 and ρ_2, we take $\rho_1 \subseteq \rho_2$ iff each element of ρ_1 is a subset of an element of ρ_2. Intuitively, ρ_1 is a finer-grained version of ρ_2 if $\rho_1 \subseteq \rho_2$, with $\rho_1(z)$ always providing at least as much information about z as does $\rho_2(z)$. So ρ_1 is a delineation of a set of degrees of freedom that includes those delineated by ρ_2. Note that $\forall\ \rho$, $\emptyset \subseteq \rho \subseteq \hat{}\emptyset$, just as in conventional set theory.

One special case of having $\rho_1 \subseteq \rho_2$ is where every element of ρ_1 occurs in ρ_2, as in the traditional notion of set inclusion. (For our purposes we can broaden that special case, which is what we've done in our definition.) Note also that the $\subseteq$ relation is transitive and that both $\rho_1 \subseteq \rho_2$ and $\rho_2 \subseteq \rho_1$ iff $\rho_1 = \rho_2$, and that $\rho_1 \subseteq \rho_2$ means there are $\hat{}\rho_1$ and $\hat{}\rho_2$ such that $\hat{}\rho_2 \subseteq \hat{}\rho_1$, just as in conventional set theory.

The other set-theory-like operations over coordinates can be defined by generalizing from the special case of conventional vector subscripts. For example, $\rho_1 \cap \rho_2$ is shorthand for a coordinate whose members are given by the intersections of the members of ρ_1 and ρ_2. We make this definition to accord with the conventional vector subscript interpretation of $z_{\rho_1 \cup \rho_2}$ as having its elements be the surfaces in ζ of both constant z_{ρ_1} and constant z_{ρ_2}. (E.g., when $\zeta = \Re^3$ and has elements written as (x, y, z), "$z_{x \cup y}$" means $z_{x,y}$, which is the set of points of constant z_x and z_y.) Given this interpretation, write $z_{\rho_1 \cup \rho_2} = \hat{}(\rho_1 \cup \rho_2) \equiv \hat{}\rho_1 \cap \hat{}\rho_2$. This then means that the elements of $\rho_1 \cap \rho_2 = z_{\hat{}\rho_1 \cup \hat{}\rho_2}$ should be surfaces of constant $z_{\hat{}\rho_1} = \rho_1(z)$ and constant $z_{\hat{}\rho_2} = \rho_2(z)$, exactly as our definition of the intersection operator stipulates.

Note that $\rho_1 \cap \rho_2 \subseteq \rho_1$, as one would like. Intuitively, the intersection operator is just the comma operator given by Cartesian products. (E.g., when $\zeta = \Re^3$ and has elements written as (x, y), $z_x \cap z_y$ is indexed by the vector (z_x, z_y).)

Finally, the intersection operator defines the union operator as $\rho_1 \cup \rho_2 = \hat{}(\hat{}\rho_1 \cap \hat{}\rho_2) = \hat{}(z_{\rho_1} \cap z_{\rho_2})$. To illustrate this, in the example of $\Re^3$, where the elements of ρ_x are lines of constant (y, z), and the elements of ρ_y are lines of constant (x, z), the elements of $\rho_x \cup \rho_y$ are planes of constant z. Similarly, when $\rho_1 \subseteq \rho_2$, $\rho_2 \backslash \rho_1$ is shorthand for a particular coordinate $\rho \subseteq \rho_2$ that is disjoint from ρ_1 (i.e., such that $\rho_1 \cap \rho = \emptyset$) and such that $\rho_1 \cup \rho = \rho_2$. Both operations are not single-valued, in general.

Note that in analogy to set theory, any coordinate ρ_1 such that there is no $\rho_2 \subsetneq \rho_1$ is equal to the null set coordinate. The analogue of a "single-element set" is a coordinate ρ that contains only itself and the null set. This is any coordinate all of whose members but one consist of a single $z \in C$, where that other member consists of two such z.

C Miscellaneous Proofs

Proof of Thm. 1: Choose any $z', z'' \in \rho(z)$. $\text{sgn}[N_{\rho,U_1}(z') - N_{\rho,U_1}(z'')] = \text{sgn}[U_1(z') - U_1(z'')]$ for all such z' and z'', by definition of intelligence. Similarly, $\text{sgn}[N_{\rho,U_2}(z') - N_{\rho,U_2}(z'')] = \text{sgn}[U_2(z') - U_2(z'')]$ for all such points. But by hypothesis, $N_{\rho,U_2}(z'') = N_{\rho,U_1}(z'')$ and $N_{\rho,U_2}(z') = N_{\rho,U_1}(z')$. So $\text{sgn}[N_{\rho,U_1}(z') - N_{\rho,U_1}(z'')] = \text{sgn}[N_{\rho,U_2}(z') - N_{\rho,U_2}(z'')]$. Transitivity then establishes the forward direction of the theorem.

To establish the reverse direction, simply note that $\text{sgn}[U_1(z') - U_1(z'')] = \text{sgn}[U_2(z') - U_2(z'')] \; \forall \; z' \in \rho(z)$, by hypothesis, and therefore by the first part of the definition of intelligence, U_1 and U_2 have the same intelligence at z''. Since this is true for all $z'' \in \rho(z)$, U_1 and U_2 have the same intelligence throughout $\rho(z)$. **QED**.

Proof of Thm. 2: Consider any $z', z'' \in \rho(z)$. We can always write $\text{sgn}[U_2(z'') - U_2(z')] = \text{sgn}[\Phi(U_2(z''), \rho(z)) - \Phi(U_2(z'), \rho(z))]$, due to the restriction on Φ. Therefore U_1 and U_2 have the same intelligence at z', by the first part of the definition of intelligence. Since this is true $\forall \; z' \in \rho(z)$, U_1 and U_2 are factored at z. This establishes the backwards direction of the proof.

For the forward direction, use Thm. 1 and the fact that the system is factored to establish that $\forall \; z$ in C, $\forall \; z'', z' \in \rho(z)$, $U_1(z') = U_1(z'')$ iff $U_2(z') = U_2(z'')$. Therefore for all points in $\rho(z)$, the value of U_1 can be written as a single-valued function of the value of U_2. Since Thm. 1 also establishes that $U_1(z') > U_1(z'')$ iff $U_2(z') > U_2(z'')$, we know that that single-valued function must be strictly increasing. Identifying that function with Φ completes the proof. **QED**.

Proof of Thm. 3: $\text{CDF}(V(\omega, k) \mid l^a, k) < \text{CDF}(V(\omega, k) \mid l^b, k)$ means that for any fixed z', with $y \equiv V(z')$,

$$P(w : V(w, k) \leq y \mid l^a, k) \;\; < \;\; P(w : V(w, k) \leq y \mid l^b, k).$$

This is equivalent to

$$P(z : V(\omega(z), \kappa(z)) \leq y \mid l^a, k) \quad < \quad P(z : V(\omega(z), \kappa(z)) \leq y \mid l^b, k),$$

i.e.,

$$P(z : V(z) \leq y \mid l^a, k) \quad < \quad P(z : V(z) \leq y \mid l^b, k).$$

Since $z \in k$ in both of these probabilities, by the second part of the definition of intelligence we get

$$P(z : N_{\kappa,V(.,k)}(z) < N_{\kappa,V(.,k)}(z') \mid l^a, k) \\ < \quad P(z : N_{\kappa,V(.,k)}(z) \leq N_{\kappa,V(.,k)}(z') \mid l^b, k) \; \forall z' \in k.$$

This in turn is equivalent to $\text{CDF}(N_{\kappa,V(.,k)} \mid l^a, k) < \text{CDF}(N_{\kappa,V(.,k)} \mid l^b, k)$.

Next write $E(N_{\kappa,V(.,k)} \mid n, k) = \int_0^1 \mathrm{d}y \; y \; P(N_{\kappa,V(.,k)} = y \mid n, k)$. Integrate by parts to get

$$E(N_{\kappa,V(.,k)} \mid l^a, k) - E(N_{\kappa,V(.,k)} \mid l^b, k) = \\ \int_0^1 \mathrm{d}y \; [\text{CDF}(N_{\kappa,V(.,k)} \mid l^b, k) - \text{CDF}(N_{\kappa,V(.,k)} \mid l^a, k)].$$

Since $\forall y, \text{CDF}(N_{\kappa,V(.,k)} \mid l^a, k)(y) < \text{CDF}(N_{\kappa,V(.,k)} \mid l^b, k)(y)$, this last integral cannot be negative. The analog for equalities of CDF's and expectations rather than inequalities follows similarly **QED**.

Proof of Lemma 1: Since both P_i are normalized and they are distinct (if they aren't distinct, we're done), $\exists u^*$ such that $P_1(u^*) > P_2(u^*)$. By our condition concerning the P_i, $P_1(u) > P_2(u) \; \forall u > u^*$. Similarly there exists a u everywhere below which P_2 exceeds P_1. Accordingly, there is a greatest lower bound on the u^*'s, T. $\forall \; y \leq T, P_1(u \leq y) \leq P_2(u \leq y)$, and therefore by the non-negativity of ϕ', $\forall \; y \leq \phi(T), P_1(u : \phi(u) \leq y) \leq P_2(u : \phi(u) \leq y)$. So the CDF of ϕ according to P_1 is less than that according to P_2 everywhere below T. Therefore if there is to be any y value at which the CDF of ϕ according to P_1 is greater than that according to P_2, there must be a least such y value, and therefore a corresponding least such u, u'. We know that $u' > T$. However for all $u > T$, $P_1(u) > P_2(u)$. Therefore $P_1(u : \phi(u) \geq \phi(u')) \geq P_2(u : \phi(u) \geq \phi(u'))$. Summing the P_1 probabilities of $\phi(u)$ exceeding and being less than $\phi(u')$, and doing the same for P_2, we see that both P_i cannot be normalized, which is impossible. QED.

Proof of Thm. 4: When the ψ's both equal $\underline{\gamma}_\rho$ and $\lambda = \nu$, by its definition H must be the actual associated n-conditioned distributions over x, $P(x \mid n^a)$ and $P(x \mid n^b)$.

To complete the proof we must demonstrate that there is at least one parametric form for H that obeys the condition in the theorem when one of the ψ's does not equal $\underline{\gamma}_\rho$ and/or $\lambda \neq \nu$. We do this by construction. First take the derivative of each ambiguity (one for each x) to get the convolutions $\int \mathrm{d}y^1 \mathrm{d}y^2 P_\psi(y^1; l, x^1) P_\psi(y^2; l, x^2) \delta(y - (y^1 - y^2))$. Multiply each such convolution by y and integrate the result over all y. This gives us the differences between the means of all the distributions $P_\psi(y; l, x)$ (one distribution for each x). Translate all those means, $M(\psi, l, x)$, by the same amount so that the lowest one has value 1. Then take $P^{[\psi;\lambda]}(x^1 \mid l) \propto e^{M(\psi,l,x)}$.

Use the relation between ordered and unordered ambiguity to rewrite the condition in the theorem as $t_U(x^1, x^2)A(\psi^a; l^a, x^1, x^2) < t_U(x^1, x^2)A(\psi^b; l^b, x^1, x^2)$. Consider some particular pair x^1, x^2, where without loss of generality $t_U(x^1, x^2) = 1$. Integrate $A(y; \psi^a; l^a, x^1, x^2) - A(y; \psi^b; l^b, x^1, x^2)$ by parts. So long as $y[A(y; \psi^a; l^a, x^1, x^2) - A(y; \psi^b; l^b, x^1, x^2)]$ goes to 0 as y goes to either positive or negative infinity, the result is

$$-[(M(\psi^a, l^a, x^1) - M(\psi^a, l^a, x^2)) - (M(\psi^b, l^b, x^1) - M(\psi^b, l^b, x^2))].$$

By hypothesis, $t_U(x^1, x^2)$ times this expression must be negative. Therefore $\frac{P^{[\psi^a;\lambda]}(x^1|l^a)}{P^{[\psi^a;\lambda]}(x^2|\lambda)} > \frac{P^{[\psi^b;\lambda]}(x^1|l^b)}{P^{[\psi^b;\lambda]}(x^2|l^b)}$. Now apply Lemma 1. **QED.**

Proof of Coroll. 2: Expand $E(U \mid r, s) = \int \mathrm{d}n\mathrm{d}x P(n \mid r, s)U(x, r)P^{[\nu]}(x \mid n)$. By the second premise we can write this integral as

$$\begin{aligned}
\int \mathrm{d}n\mathrm{d}x P(n \mid r, s)U(x, r)P^{[\nu,\sigma]}(x \mid n, s) &= \int \mathrm{d}n\mathrm{d}x P(n \mid r, s)U(x, r)P^{[\underline{g}_s;\nu,\sigma]}(x \mid n, s) \\
&= \int \mathrm{d}n\mathrm{d}x P(n \mid r, s)U(x, r)P^{[\underline{g}_s;\nu,\sigma]}(x \mid n, s, W) \\
&= \int \mathrm{d}n P(n \mid r, s)E^{[\underline{g}_s;\nu,\sigma]}(U \mid n, s, W).
\end{aligned}$$

QED.

Proof of Coroll. 3: For both $\psi = \underline{g}_{s^a}$ and $\psi = \underline{g}_{s^b}$, expand

$$E^{[\psi;\nu,\sigma]}(N_\rho \mid n, s^b) = \int \mathrm{d}r \frac{P(n \mid r, s^b)P(r \mid s^b)}{P(n \mid s^b)} E^{[\psi;\nu,\sigma]}(N_\rho \mid n, r, s^b).$$

Rearranging terms gives the hypothesis inequality of our corollary. Now apply Coroll. 2 to the consequent inequality of the third premise with $\Omega = \eta = \emptyset$. **QED.**

Proof of Thm. 5: By condition (iv), the quantity y^* defined there must equal $\underline{g}_{s^a}(x^*, r)$. Now fix x^1 and x^2. By conditions (ii) and (iii), for both of those moves x^i, $\underline{g}_{s^a}(x^i, r')$ has either the value 0 or 1 for all r' arising in the expansion of $A(\underline{g}_{s^a}; n, x^1, x^2)$. Combining this with the value of y^*, we see that for any r and any pair (x^1, x^2), one of the following four cases must hold:

I) $\underline{g}_{s^a}(x^1, r) = 0$, and
$P(\underline{g}_{s^a} = y; n, x^1)$ is a delta function about 0, and
$P(\underline{g}_{s^a} = y; n, x^2)$ is an average of two delta functions, centered about 0 and about 1;

II) $\underline{g}_{s^a}(x^1, r) = 1$, and
$P(\underline{g}_{s^a} = y; n, x^1)$ is a delta function about 1, and
$P(\underline{g}_{s^a} = y; n, x^2)$ is an average of two delta functions, centered about 0 and about 1.

(Cases (III) and (IV) are the same as (I) and (II), just with x^1 and x^2 interchanged.)

Without loss of generality assume that we're in case (II). Then expand $A(y; U, \underline{g}_{s^a}; n, x^1, x^2)$ as

$$\int dy^1\, dy^2 P(\underline{g}_{s^a} = y^1; n, x^1) P(\underline{g}_{s^a} = y^2; n, x^2)\Theta[y - (y^1 - y^2)\,\mathrm{sgn}[U(x^1, r) - U(x^2, r)]]. \tag{C.2}$$

This evaluates as

$$\int dy^2\, P(\underline{g}_{s^a} = y^2; n, x^2)\Theta[y - (1 - y^2)\,\mathrm{sgn}[U(x^1, r) - U(x^2, r)]].$$

Now $\mathrm{sgn}[\underline{g}_{s^a}(x^1, r) - \underline{g}_{s^a}(x^2, r)]$ equals 0 or 1 for case (II). So by condition (i), and the factoredness of $\underline{g}_{s^a}$ and $\underline{g}_{s^b}$, this must also be true for $\mathrm{sgn}[U(x^1, r) - U(x^2, r)]$. Given that y^2 cannot exceed 1, this in turn means that the theta function is nonzero only for non-negative y. Accordingly, so is the ambiguity.

This character of the ambiguity holds for all four cases; for all of them the ambiguity $A(y; \underline{g}_{s^a}, n, x^1, x^2)$ is 0 up to $y = 0$ where it may have a jump, and then is flat up to 1, where if the first jump did not go up to 1 it now has a second jump that gets it up to 1. So its support is assuredly non-negative. **QED.**

Proof of Thm. 6: Define $m \equiv t_U(x^1, x^2)$. Our condition means that

$$\int dy^1\, dy^2\, \Theta[y - (y^1 - y^2)m] P(V_a(x^1, \rho) = y^1 \mid l') P(V_a(x^2, \rho) = y^2 \mid l') \quad =$$
$$\int dy^1\, dy^2\, \Theta[y - (y^1 - y^2)m] \;\; P(KV_b(x^1, \rho) + h(x^1) = y^1 \mid l) P(KV_b(x^2, \rho) + h(x^2) = y^2 \mid l),$$

i.e.,

$$\int dy^1\, dy^2\, \Theta[y - (y^1 - y^2)m][P(V_a(x^1, \rho) = y^1 \mid l') P(V_a(x^2, \rho) = y^2 \mid l')]$$

$$= \int dr^1\, dr^2 \Theta[y - mK(V_b(x^1, r^1) - V_b(x^2, r^2)) - K(h(x^1) - h(x^2))] P(r^1, r^2; l)$$
$$= \int dr^1\, dr^2\, \Theta[y/K - m(V_b(x^1, r^1) - V_b(x^2, r^2)) - m(h(x^1) - h(x^2))/K] P(r^1, r^2; l)$$
$$= \int dy^1\, dy^2\, \Theta[\{y/K - m(h(x^1) - h(x^2))/K\} - (y^1 - y^2)m]$$
$$P(V_b(x^1, \rho) = y^1 \mid l) P(V_b(x^2, \rho) = y^2 \mid l).$$

QED.

Proof of Thm. 7: To prove (i), first marginalize out y^2 from the equality relating PV_a and P_{KV_b+h}, and then use the resultant equality between probability distributions to form an equality concerning the two associated variances of y^1. The resultant formula for K holds for any x^1, and therefore it holds under arbitrary averaging over the x^1.

To prove (ii), use the equality relating P_{V_a} and P_{KV_b+h} to relate the expected values of the difference $(y^1 - y^2)$, evaluated according to the two distributions P_{V_a} and P_{V_b}:

$$\int \mathrm{d}r^1 \, \mathrm{d}r^2 \, P(r^1, r^2; l', x^1, x^2)[V_a(x^1, r^1) - V_a(x^2, r^2)]$$
$$= h(x^1) - h(x^2) + K \int \mathrm{d}r^1 \, \mathrm{d}r^2 \, P(r^1, r^2; l', x^1, x^2)[V_b(x^1, r^1) - V_b(x^2, r^2)].$$

Next collect terms to get an expression for $[h(x^2) - h(x^1)]/K$ in terms of expected values of V_a and V_b. Finally plug in the definition of Λ_f and evaluate K to verify our equation for $[h(x^2) - h(x^1)]/K$. **QED.**

Proof of Thm. 8: To prove (i), note that since $P(r'; l) = P(r'; l')$, and since V_a and V_b have the same lead utility, $E(V_a; l', x^1) - E(V_a; l', x^2) = E(V_b; l, x^1) - E(V_b; l, x^2)$. Therefore the drop in learnability means that $\int \mathrm{d}x \, f(x) \, \mathrm{Var}(V_a; l', x) < \int \mathrm{d}x \, f(x) \, \mathrm{Var}(V_b; l, x)$. Plugging this into Thm. 7(i) gives the result claimed.

To prove the second part of (ii), for pedagogical clarity define $m \equiv t_{V_A}(x^1, x^2)$ and write the derivative as

$$\int \mathrm{d}r^1 \, \mathrm{d}r^2 \, P(r^1, r^2; l', x^1, x^2)\delta(m[V_a(x^1, r^1) - V_a(x^2, r^2)])$$
$$= \int \mathrm{d}r^1 \, \mathrm{d}r^2 \, P(r^1, r^2; l, x^1, x^2)\delta(m[K\{V_b(x^1, r^1) - V_b(x^2, r^2)\} + h(x^1) - h(x^2)])$$
$$= K^{-1} \int \mathrm{d}r^1 \, \mathrm{d}r^2 \, P(r^1, r^2; l, x^1, x^2)$$
$$\delta(m[V_b(x^1, r^1) - V_b(x^2, r^2)] - \frac{m[\Lambda_f(V_{\ b}; l, x^1, x^2) - \Lambda_f(V_a; l', x^1, x^2)]}{\sqrt{\int \mathrm{d}x f(x) Var(V_b; l, x^1, x^2)}}),$$

where Thm. 7(ii) was used in the last step. By hypothesis, the difference in learnabilities equals zero though. This establishes the result claimed.

To prove the first part of (ii), use similar reasoning to write the value of the ambiguity at the origin as

$$\int \mathrm{d}r^1 \, \mathrm{d}r^2 \, P(r^1, r^2; l', x^1, x^2)\Theta(m[V_a(x^1, r^1) - V_a(x^2, r^2)])$$
$$= \int \mathrm{d}r^1 \, \mathrm{d}r^2 \, P(r^1, r^2; l, x^1, x^2)$$
$$\delta(m[V_b(x^1, r^1) - V_b(x^2, r^2)] - \frac{m[\Lambda_f(V_{\ b}; l, x^1, x^2) - \Lambda_f(V_a; l', x^1, x^2)]}{\sqrt{\int \mathrm{d}x f(x) Var(V_b; l, x^1, x^2)}}).$$

(iii) is immediate from Thm. 7(i).

Finally, to prove (iv), without loss of generality take $K < 1$, and use the trick in (ii) with $s^* = s$ to increase K to 1. Doing this reduces the maximal slope of the associated ambiguity. In addition, it results in a right-shifted version of the ambiguity $A(V_b; l, x^1, x^2)$. Therefore this reduced maximal slope is the same as the maximal slope of $A(V_b; l, x^1, x^2)$. **QED.**

Proof of Coroll. 5: Due to their all obeying Coroll. 4(ii), all utilities share the same m, which equals all of their m''s. Write

$$\begin{aligned} A(y; V^*(., r), V^*; l_{t^*}, x^1, x^2) & \\ = & \int \mathrm{d}r^1\, \mathrm{d}r^2\, P(r^1, r^2; l_{t^*}, x^1, x^2)\Theta[y - m(V^*(x^1, r^1) - V^*(x^2, r^2))] \\ = & \int \mathrm{d}r^1\, \mathrm{d}r^2\, P(r^1, r^2; l_t, x^1, x^2) \ \times \\ & \qquad \Theta[(\{y/K_{l_{t^*}, V^*, l_t, V_t}\} - \Delta_{t,t^*,x^1,x^2}) - m(V_t(x^1, r^1) - V_t(x^2, r^2))]. \end{aligned}$$

On the other hand,

$$A(y; V_t(., r), V_t; l_t, x^1, x^2) = \int \mathrm{d}r^1\, \mathrm{d}r^2\, P(r^1, r^2; l_t, x^1, x^2)\Theta[y - m(V_t(x^1, r^1) - V_t(x^2, r^2))].$$

By comparing our formulas for the two ambiguities, we see that as long as

$$\frac{y}{K_{t,t^*}} - \Delta_{t,t^*,x^1,x^2} \leq y \ \ \forall \ \ y \in [D_{t,x^1,x^2}, B_{t,x^1,x^2}],$$

it follows that $A(V_t(., r), V_t; l_t, x^1, x^2) \geq A(V^*(., r), V^*; l_{t^*}, x^1, x^2)$. Furthermore, by our formulas for algebraic manipulation of K's, we know that $K_{l_{t^*}, \beta K_{t,t^*}, l_t, V_t} = K_{l_{t^*}, \beta K_{t,t^*}, l_{t^*}, V^*} K_{l_{t^*}, V^*, l_t, V_t}$. By Thm. 8(iii), this just equal $\beta K_{l_{t^*}, V^*, l_t, V_t} = \beta K_{t,t^*}$.

Accordingly, L_{t,t^*,V^*,x^1,x^2} is the set of values β by which one could multiply K_{t,t^*} and still have the desired inequality hold, given the values of D_{t,x^1,x^2} and B_{t,x^1,x^2}. L_{t,t^*,V^*} is then defined as the set of such multiples for which we can be assured that the inequality holds for every (x^1, x^2) pair. So for every β in that set, we know that $(\beta V^*, l_{t^*})$ has better ambiguity than does (V_t, l_t), for every single (x^1, x^2) pair. Accordingly, by Coroll. 1, it has better expected intelligence as well. That means that so long as $\beta \in \cup_t L_{t,t^*,V^*}$, it follows that $(\beta V^*, l_{t^*})$ has better expected intelligence than *some* (V_t, l_t). **QED.**

Proof of Thm. 9: By Thm. 2, a utility U_1 is factored with respect to U_2 for agent ρ at z iff we can write it as $U_1(z') = \Phi_r(U_2(z'))$ for some r-parameterized function Φ whose first partial derivative is positive across all $z' \in \rho(z)$. Any such function can always be written as $F_r(U_2) - D$ for some function D only dependent on $\rho(z)$ and some f-parameterized functionF_r whose derivative is positive. This establishes (i).

To minimize the learnability of U_1 given Φ, l, and U_2, first note that since D is independent of x, the numerator in the definition of $\Lambda_f(U_1; l, x^1, x^2)$, $E(U_1; l, x^1) - E(U_1; l, x^2)$, is independent of the choice of D. So we need only consider the denominator. Rewrite that denominator as

$$E_{f(x)}[\mathrm{Var}(U_1; l, \xi)] \\ = (1/2) \int \mathrm{d}x\, f(x) \int \mathrm{d}r'\, \mathrm{d}r''\, P(r'; l) P(r''; l) [U_1(x, r') - U_1(x, r'')]^2$$

where we have used the fact that $\mathrm{Var}_{\{A(\tau\}} = (1/2) \int \mathrm{d}t_1\, \mathrm{d}t_2\, P(t_1) P(t_2) [A(t_1) - A(t_2)]^2$ for any random variable τ with distribution P.

Bring the integral over x inside the other integrals, expand U_1, and introduce the shorthand $D_1(x, r) \equiv F_r(U_2(x, r)$ to get

$$(1/2) \int \mathrm{d}r'\, \mathrm{d}r''\, P(r'; l) P(r''; l) \int \mathrm{d}x\, f(x) [D_1(x, r') - D_1(x, r'') - (D(r') - D(r''))]^2.$$

The innermost integral is minimized for each r', and r'' so long as for each r' and r'',

$$D(r') - D(r'') = \int \mathrm{d}x\, f(x) [D_1(x, r') - D_1(x, r'')].$$

This can be assured by picking $D(r) = E_{f(x)}(D_1(\xi, r))$ for all r. This establishes (ii).

Since $E(U_1; r, s, x^1) - E(U_1; r, s, x^2) = E(U_2; r, s, x^1) - E(U_2; r, s, x^2)$, the ambiguity shift in going from U_2 to U_1 equals

$$(E(U_1; l, x^1) - E(U_1; l, x^2)) \left\{ 1 - \sqrt{\frac{E_f(\mathrm{Var}(U_2; l, \xi))}{E_f(\mathrm{Var}(U_1; l, \xi))}} \right\}.$$

So what we need to do is minimize $\frac{E_f(\mathrm{Var}(U_2; l, \xi))}{E_f(\mathrm{Var}(U_1; l, \xi))}$.

Now for our choice of D, by the reasoning above,

$$E_f(\mathrm{Var}(U_1; l, \xi)) = (1/2) \int \mathrm{d}r'\, \mathrm{d}r''\, P(r'; l,) P(r''; l)\, \mathrm{Var}_{f(x)}(D_1(\xi, r') - D_1(\xi, r'')).$$

Now again use the fact that $\mathrm{Var}_{\{A(\tau\}} = (1/2) \int \mathrm{d}t_1\, \mathrm{d}t_2\, P(t_1) P(t_2) [A(t_1) - A(t_2)]^2$ for any random variable τ with distribution P and associated function A to expand the Var_f into a double integral. Next rearrange terms, and again use that fact, this time to reduce the integral over r' and r'' into a single variance. **QED.**

Proof of Thm. 10: Any change to z that doesn't move it out of the set $B \cap \pi'(z)$ doesn't move it out of $B \cap \pi(z)$, since all z in any element of π' lie in the same element of π. Therefore that change to z doesn't change $\pi(z)$. That means in turn that it does not change $D_1(\mathrm{CL}_{\hat{\pi}}(z).)$ So $D_1(\mathrm{CL}_{\hat{\pi}}(z).)$ can be written as a function that depends only on $B \cap \pi'(z)$. Therefore it is of the form for the secondary utility required for the difference utility to be factored with respect to agent $B \cap \pi'(z)$. **QED.**

Proof of Thm. 11: Note that $H(\mathrm{CL}_{\neg\kappa}(z))$ can be written as a function of $\kappa(z)$, and therefore of $\rho(z)$. Accordingly, expand the numerator term in the definition of learnability in terms of r to see that that it has the same value for H and $\mathrm{WLU}_{H,\kappa}$.

Write out $\mathrm{WLU}_{H,\kappa}(z) = H_1(z) - H_1(\mathrm{CL}_{\neg\kappa}(z))$ to see that the denominator term for $\Lambda_f(\mathrm{WLU}_{H,\kappa}; l, x^1, x^2)$ is bounded above by

$$\int \mathrm{d}x\, f(x) \int \mathrm{d}r'\, P(r';l)[H_1(x,r') - H_1(\mathrm{CL}_{\neg\kappa}(x,r'))]^2.$$

In turn, the greatest possible value of the term in square brackets is M. So that denominator term is bounded above by M.

Write the denominator term for $\Lambda_f(H; l, x^1, x^2)$ as

$$\begin{aligned}
&(1/2) \int \mathrm{d}x\, f(x) \int \mathrm{d}r'\, \mathrm{d}r''\, P(r';l)P(r'';l) \times \\
&\qquad\qquad [\{H_2(\kappa(r')) - H_2(\kappa(r''))\} + \{H_1(x,r') - H_1(x,r'')\}]^2 \\
&= (1/2) \int \mathrm{d}x\, f(x) \int \mathrm{d}r'\, \mathrm{d}r''\, P(r';l)P(r'';l) \\
&\qquad\qquad \{[H_2(\kappa(r')) - H_2(\kappa(r''))]^2 + [H_1(x,r') - H_1(x,r'')]^2 + \\
&\qquad\qquad\qquad 2[H_2(\kappa(r')) - H_2(\kappa(r''))][H_1(x,r') - H_1(x,r'')]\}.
\end{aligned}$$

The third of the integrals summed in this last expression is bounded below by

$$-\sqrt{M} \int \mathrm{d}x\, f(x) \int \mathrm{d}r'\, \mathrm{d}r''\, P(r';l)P(r'';l)|H_2(\kappa(r')) - H_2(\kappa(r''))|,$$

which in turn is bounded below by $-\sqrt{ML}$, due to concavity of the squaring operator. The second of our integrals is bounded below by 0. Finally, the first of these integrals equals $L/2$ exactly. Combining, the denominator term for $\Lambda_f(H; l, x^1, x^2)$ is bounded below by $L/2 - \sqrt{ML}$. **QED.**

D Repeating Coordinates, Multi-Step Games, and Constrained Optimization

Say we have a set of coordinates of ζ, indicated by $\{\zeta^1, \zeta^2, \ldots, \zeta^T\}$, with associated images of C written as $\{C^1, C^2, \ldots, C^T\}$. Conventionally the index t is called "time" or the "timestep". An associated **repeating coordinate** is a set $\{\lambda^1, \lambda^2, \ldots, \lambda^T\}$ such that $\forall\, t$, $\lambda^t(z) = \lambda(\zeta^t(z))$ for some function λ whose domain is given by the union of the ranges of the coordinates $\{\zeta^i\}, Z$. For a **deterministic** set $\{\zeta^i\}$, there is a set of single-valued functions $\{E^i\}$, mapping Z to Z, such that $\zeta^{i+1} = E^i(\zeta^i)\ \forall\, i \in \{1, \ldots, T-1\}$. The set is **time-translation-invariant** if E^i is the same for all i, and **(temporally) invertible** if the E^i are all invertible.

In close analogy to conventional game theory nomenclature, we say that we have a set of **players** $\{i\}$, each consisting of a separate triple of repeating coordinates $\{\rho_i^t\}$, $\{\xi_i^t\}$, and $\{\nu_i^t\}$, if for each t and i the triple $(\rho_i^t, \xi_i^t, \nu_i^t)$ act as the context, move, and worldview coordinates, respectively, of an agent. If in addition $T > 1$, we sometimes say we have a **multi-step game**, and identify each "step" with a different time.

Often we want to consider the intelligences of the players' agents with respect to some associated sequences of private utilities. We can do this if in addition to the players we have a repeating coordinate $\{\sigma^t\}, s^1$ being the design coordinate value set by the designer of the collective, and $\underline{g}_{i^t}(z) \equiv \underline{g}_{i,\sigma^t(z)}(z)$ being the private utility of player i at time t.[55] In this way each player is identified with a sequence of agents.

A **multi-stage game** is one in which for every i, $\underline{g}_{i^t}$ is the same function of $z^T \in Z$. A **normal-form** (version of a multi-stage) **game** is the system ζ^1 with associated coordinates and set of allowed points C^1, where $P(z_1)$ is set by marginalizing $P(z)$. So in particular, $P(\underline{g}_{i^1}(z_1) = v) = \int \mathrm{d}z\, P(z^T \mid z_1)\delta(v - \underline{g}_{i^T}(z^T))$. Intuitively, a normal form game is the underlying multi-stage game "rolled up" into a single stage, that stage being set by the initial joint state of the players.

If for every i, $\underline{g}_{i^t}$ is the same function from $z^t \in Z$ to the reals, then we say we have an **iterated game**. More generally, if for each player i all of the $\{\underline{g}_{i^t}\}$ are the same discounted sum over $t' \in \{1, \ldots, T\}$ of $R_i(z^{t'})$ for some real-valued **reward** function R_i that has domain Z, then each player's agents must try to predict the future, and we have a **repeated game**.

Note that conventional full rationality noncooperative game theory of normal form games, involving Nash equilibria of the private utilities, is simply the analysis of scenarios in which the intelligence of z with respect to each player's private utility, given the context set by the other players' moves, equals 1. This fact suggests many extensions of conventional noncooperative game theory based on the formalism of this paper. For example, we can consider games in which $C \neq \zeta$, i.e., not all joint-moves are possible. Another modification, applicable if we use the percentile-type of intelligence, is to restrict $\mathrm{d}\mu_\rho$ to some limited "set of moves that player ρ actively considers". This provides us with the concept of an "effective Nash equilibrium" at the point z, in the sense that *over the set of moves it has considered*, each player has played a best possible move at such a point. In particular, for moves in a metric space, we could restrict each $\mathrm{d}\mu_\rho$

[55] An interesting topic is whether for a particular player there is a set of functions $\{U^t(z^t)\}$ such that the values $\{x^t\}$ induce large $N_{\rho^t, U^t}(z^t)$, $\forall\, t \in \{1, \ldots, T\}$. When there is such a set, it would seem natural to interpret the player as a set of "agents" with associated private utilities $\{U^t\}$. However unless we can vary the private utility that the time t "agent" is supposedly trying to maximize, we have no reason to believe that the value x^t really is set by a learning algorithm trying to maximize that private utility. (We might have a coordinate akin to the explicitly non-learning spins in Ex. 1 of [22].) This means that for such an interpretation be tested, the-private utility must be part of some $\{\sigma^t\}$, so we can set it. Our modifying it must then induce associated changes in the moves consistent with the supposition that a learning algorithm is controlling those moves to try to maximize those values of the private utilities, as discussed in the subsection on the first premise.

to some infinitesimal neighborhood about z, and thereby define a "local Nash equilibrium" by having ρ's intelligence with respect to utility $\underline{\gamma}_\rho$ equal 1 for each player ρ.

More generally, as an alternative to fully rational games, one can define a bounded rational game as one in which the intelligences equal some vector $\vec{\varepsilon}$ whose components need not all equal 1. Many of the theorems of conventional game theory can be directly carried over to such bounded-rational games [19] by redefining the utility functions of the players. In other words, much of conventional full rationality game theory applies even to games with bounded rationality, under the appropriate transformation. This result has strong implications for the legitimacy of the common criticism of modern economic theory that its assumption of full rationality does not hold in the real world, implications that extend significantly beyond the Sonnenschein-Mantel-Debreu Theorem equilibrium aggregate demand theorem [11].

Note also that at any point z that is a Nash equilibrium in the set of the player's utilities, every player's intelligence with respect to its utility must equal 1. Since that is the maximal value any intelligence can take on, a Nash equilibrium in those utilities is a Pareto optimal point in the values of the associated intelligences (for the simple reason that no deviation from such a z can raise any of the intelligences). Conversely, if there exists at least one Nash equilibrium in the player utilities, then there is not a Pareto optimal point in the values of the associated intelligences that is not a Nash equilibrium.

Note that the moves of some player i may directly set the private utility functions of the agent(s) of some other player i' in a multi-step game. In particular, the private utilities of i's agents might explicitly involve inferences about the effect on $P(G \mid s^t)$ of various possible choices of $\underline{g}_{(i')^t}$. Loosely speaking, when an agent of player i changes the learning algorithm, move variable, world-view variable, and/or private utilities of (the agents of) other players, and does so gradually, based on considerations of how to improve $P(G \mid s^t)$, we refer to its learning algorithm as engaging in **macrolearning**; that agent's moves constitute on-line modification of s to try to improve G. We contrast this with **microlearning**, in which one agent's moves are not viewed as directly setting other agents' private utility functions, in loose analogy with the distinction between macroeconomics and microeconomics.[56]

In any kind of game, each agent only works to (try to) maximize its current private utility.[57] However $\underline{g}_{i^t}$ will not be mutually factored (with respect to moves x^t) with either the utilities $\underline{g}_{i^{t'\neq t}}$ or with G, in general. Intuitively, moves that improve the current private utility may hurt the future one, and may even

[56] In general, we wish to optimize G *subject to the communication restrictions at hand.* When the nodes are agents, such restrictions apply to the argument lists of their private utilities. More generally though, the nodes can communicate with each other in ways other than via their private utilities. Indeed, part of macrolearning in the broadest sense of the term is modifying such extra-utility "signaling" and "bargaining" among the nodes, to try to improve performance of the overall system. None of these "low level" issues are addressed in this paper.

[57] Formally, the first premise applies to moves and private utilities that share the same time, since here the full agent is defined for a single time.

(due to those future effects) hurt G. (See [1] for an example of this). In repeated games where G is itself a discounted sum, appropriate coupling of the reward function of the player with that of G can ensure factoredness of those two reward functions. However in iterated games—which for example are those that arise with the Boltzmann learning algorithms considered in [17]—there is no such assurance. And even for repeated games with discounted sum G's, simply having each of the player's rewards be factored with respect to the associated reward of G does not ensure that the player's full private utility is factored with respect to G.[58]

Another subtlety arises if there is randomness in the dynamics of the system at times $t' > t$, and we are considering a utility function at time t that depends on components of z other than z^t (e.g., we have a multi-stage game). The problem is that in general we require utility functions to be expressible as a single-valued function of the move and context of any agent. So in particular our utility must be such a function of (x_i^t, r_i^t), despite the stochasticity at times $t' > t$.

One way around this problem is not to cast the problem as a multi-step game, and instead have contexts explicitly includes future states of the system. We can keep the game-theoretic structure though if we have z specify the state of the pseudo-random number generator underlying the stochasticity, and then have that state be included in r_i^t. This encapsulates the stochastic dynamics within a deterministic system. Another approach is to recast utilities and associated intelligences in terms of partial worldpoints $z^{t' \leq t}$ rather than full worldpoints that include time to the future of t. As an example, starting with a conventional utility U, we could define a new utility $\hat{U}(z) \equiv E(U \mid z^{t' \leq t})$. Since $\hat{U}(z') = \hat{U}(z)$ if $(z')^{t' \leq t} = z^{t' \leq t}$, $N_{\rho, \hat{U}}(z)$ only judges z by the quality of its components for times previous to the future.

There is another subtlety that can arise even in deterministic games, from the general requirement that any move can accompany any context. The problem is that this requirement is, on the face of it, incompatible with constrained optimization problems, in which typically for any moment t C forbids some of the potential joint-states of the agents at that time. The simplest way around this difficulty, when it is feasible, is simply to choose a different set of move coordinates for the agents, one in which the constraints do not restrict the agent's moves. Another way around this difficulty is to transform the problem by means of a function that maps any (unconstrained) pair (x, r) to an allowed (constrained) joint-state of all agents, which in turn is what is used to determine

[58] In practice factoredness of reward functions often results in approximate factoredness of associated utilities if t is large enough so that the system has started to settle toward a Nash equilibrium among the players' reward functions. In turn, such settling toward a Nash equilibrium is expedited if we set s to give a good term 3 in the "reward utility version" of the central equation, in which all utilities are replaced by the associated reward functions.

For the more general scenario where factoredness of reward functions does not suffice, one can guarantee factoredness of the utilities by using reward functions set via "effect sets". As discussed in the discussion of the WLU, such reward functions can ensure factoredness by (in essence) overcompensating for all possible future effects on G of a player's current action. A more nuanced approach is investigated in [20].

utility values.

No such function is needed however if the constrained optimization problem can be cast as traversing the nodes in a graph with fixed fan-out, so that the constraints don't apply to the moves directly. To see this, first consider an iterated game with an "environment" repeating coordinate $\{\theta^t\}$. Say that the game is a Markovian control problem with N players, i.e., a multi-stage game where $G(z)$ only depends on the value q^T and

$$\begin{aligned} P(q^t \mid q^{t-1}, x_1^1, x_1^2, \ldots, x_1^{t-1}, x_2^1, \ldots, x_N^{t-1}) &= P(q^t \mid q^{t-1}, x_1^{t-1}, x_2^{t-1}, \ldots, x_N^{t-1}) \\ &= v(q^t, x_1^{t-1}, x_2^{t-1}, \ldots, x_N^{t-1}) \end{aligned}$$

where v is independent of $t \in \{1, \ldots, T-1\}$.[59]

For a graph-traversal version of this problem the dynamics is single-valued, so we can write $v(q', q, x^1, \ldots, x^N) = \delta(q' - x^1 x^2 \ldots x^N(q))$ for some function of q and $(x^1, \ldots, x^N)$ that is written as $x^1 x^2 \ldots x^N(q)$. (For uncountable q, this is a continuum-limit graph.) So any constraints o on optimizing G—on finding the optimal node q in the graph—are reflected in the graph's topology.

This kind of problem is a (fixed fan-out) undirected-graph-traversal problem if in addition the values of each ξ_i form a group, in the following sense:

i) $\forall\, q \in \theta$, $\exists!(I_1, I_2, \ldots, I_N) \in \{(x^1, x^2, \ldots x^N)\}$ such that $I_1 I_2 \ldots I_N(q) = q$;

ii) $\forall\, q \in \theta$, $\forall\, (x'_1, x'_2, \ldots x'_N) \in \{(x^1, x^2, \ldots x^N)\}$, $\exists!((x')_1^{-1}, (x')_2^{-1}, \ldots, (x')_N^{-1}) \in \{(x^1, x^2, \ldots x^N)\}$ such that $(x')_1^{-1}(x')_2^{-1} \ldots (x')_N^{-1} x'_1 x'_2 \ldots x'_N(q) = q$.

In practice, search across such a graph is easiest when the identity and inverse elements of each group of moves are independent of q, and G does not vary too quickly as one traverses the graph.

Finally, as an illustration of off-equilibrium benefits of factoredness, consider the case where ζ is a Euclidean space with an iterated game structure where every $\rho^t(z)$ is a manifold and all of those manifolds are mutually orthogonal everywhere on C. Presume that all utilities are analytic. Then for small enough step sizes, having each player run a gradient ascent on its reward function must result in an increase in G, for a factored system. (However such a gradient ascent may progressively *decrease* the values of some players' utilities.)

To see why G must increase under gradient ascent, first, as a notational matter, when M is a manifold embedded in ζ define $\nabla_M F(z)$ to be the gradient of F in some coordinate system for M, expressed as a vector in ζ. Let $\mathfrak{J}_{\rho^t}$ be the tangent plane to $\rho^t(z)$ at z. Then if G is factored with respect to $\underline{g}_{\rho^t}$, $\nabla_{\mathfrak{J}_\rho}(\underline{g}_{\rho^t}(z))$ must be parallel to $\nabla_{\mathfrak{J}_{\rho^t}}(G(z))$. (If there were any discrepancy between the directions of those two gradients, there would be a direction within $\rho^t(z)$ in which one could move z and in so doing end up increasing $\underline{g}_{\rho^t}$ but decreasing G.) So the dot product between those gradients is non-negative, and therefore changing

[59] Note that in this problem, G is not a direct function of the players' joint-move at any time. Rather the joint-move specifies the incremental change to another variable—the environment—which is what directly sets the value of G. See App. E on gradient ascent over categorical variables.

$z \rightarrow z + |\alpha| \nabla_{\mathfrak{I}_{\rho^t}}(\underline{g}_{\rho^t}(z))$ for infinitesimal α cannot decrease $G(z)$. Generalizing, note that for any utility U the gradients $\nabla_{\mathfrak{I}_{\rho^t}}(U)$ (one for each ρ^t) are mutually orthogonal, since the underlying manifolds are. Therefore having all those dot products be non-negative means that moving z an infinitesimal amount in ζ in the direction with components in each plane $\mathfrak{I}_{\rho^t}$ given by $\nabla_{\mathfrak{I}_{\rho^t}}(\underline{g}_{\rho^t}(z))$, cannot decrease $G(z)$. So gradient ascent works for factored systems.

Similarly, fix t, and consider two worldpoints z' and z'' that are infinitesimally close, but potentially differ for every player. Then it may be that for no player ρ does $\rho^t(z') = \rho^t(z'')$; every player sees a different set of the moves of its opponents at z' and z''. Nonetheless, again using non-negativity of the dot products, the system's being factored means that there must be at least one player ρ for which $\text{sgn}[G^t(z') - G^t(z'')] = \text{sgn}[g_{\rho^t}(z') - g_{\rho^t}(z'')]$. (Compare to Thm. 1.)

E Example — gradient ascent for categorical variables

This example illustrates the many connections between traditional search techniques like gradient ascent and simulated annealing on the one hand, and the use of a collective of agents to maximize a world utility on the other.

Say we have a Cartesian product space $M \equiv M^1 \times M^2 \times \cdots M^L$, where each M^i is a space of $|M^i|$ categorical (i.e., symbolic, non-numeric) variables. Write a generic element of M as m, having components $m_i, i \in \{1, \ldots, L\}$. Consider a function $h(m) \rightarrow \Re$ that we want to maximize. Because M is not a Euclidean space, we cannot use conventional gradient ascent to do this. However we can still use gradient ascent if we transform to a probability space.

To see how, take ζ to be the space of Euclidean vectors comprising the Cartesian product $S^{|M^1|} \times S^{|M^2|} \times \cdots S^{|M^L|}$, where each $S^{|M^i|}$ is the M^i-dimensional unit simplex. Define the function $R(z) \equiv \sum_{m \in M} (\prod_{i=1}^{L} z_{i,m_i}) \times h(m)$. The product $z_P \equiv (\prod_{i=1}^{L} z_{i,m_i})$ gives a (product) probability distribution over the space of possible $m \in M$. (Intuitively, $z_{i,j} = P(m^i = j)$.) Accordingly, $R(z)$ is the expected value of h, evaluated according to the distribution z_P.

Define $m^* \equiv \text{argmax}_m\, h(m)$. Then

$$\begin{aligned} \text{argmax}_z R(z) = \; [\quad & \delta(z_{1,1} - 0), \delta(z_{1,2} - 0), \ldots, \delta(z_{1,m_1^*} - 1), \ldots, \delta(z_{1,|M^1|} - 0); \\ & \ldots,\ \delta(z_{2,m_2^*} - 1), \ldots; \\ & \quad \ldots \\ & \ldots,\ \delta(z_{L,m_L^*} - 1), \ldots \quad], \end{aligned}$$

i.e., the z that maximizes R, z^*, is a Kronecker delta function about the m that maximizes h. However unlike m, z lives in (a subset of) a Euclidean space. So if we make sure to always project $\nabla R(z)$ onto $S^{|M^1|} \times S^{|M^2|} \times \cdots S^{|M^L|}$, the space of allowed z, we can use gradient ascent over z values to climb G— and thereby maximize h. Intuitively, as opposed to conventional gradient ascent over the variable of direct interest (something that is meaningless for categorical variables), here we are performing gradient ascent over an auxiliary variable, and in that way maximizing the function of the variable of direct interest.[60]

Note that R is a multilinear function over the (sub)vector spaces $\{S^{|M^i|}\}$, and its maximum must lie at a vertex of that space. There are $|M^i|$ components of the gradient of R for each variable i, giving $\sum_{i=1}^{L} |M^i|$ components altogether. The value of the component corresponding to the j'th possible value of M^i is given by the expected value of h conditioned on $m_i = j$. So calculating $\nabla R(z)$ means calculating $\sum_{i=1}^{L} |M^i|$ separate expectation values. Furthermore, at z^*, every component of the gradient has the same value, namely $h(m)$, and at all other z the value of every component of the gradient is bounded above by $h(m)$.[61]

Unfortunately, calculating $\nabla R(z)$ exactly is prohibitively difficult for large spaces. However we can readily estimate the components of the gradient instead by recasting it as a technique for improving world utility in a collective. Define $G(z \in \zeta) \equiv R(z^T \in \zeta^T)$, where z is the history of joint states of a set of agents over a sequence of T steps in an iterated game, z^t being the state at step t of the game (see App. D). Define z_i^t as the vector given by projecting z^t onto the i'th simplex $S^{|M^i|}$, i.e., the time t-value of the vector $(z_{i,1}, z_{i,2}, \ldots, z_{i,|M^i|})$. Have all LT of the Cartesian product variables $z_1^t \times z_2^t \times \cdots z_{i-1}^t \times z_{i+1}^t \times \cdots z_L^t$ be (the value of) a generalized agent coordinate ρ_i^t, $x_i^t = z_i^t$ being the value of the associated move. So for every agent, G is a single-valued function of that agent's move and its context, as required.[62]

The dynamical restrictions coupling all these distributions gives us C. To design that dynamics, note that even though $R(z^t)$ is in no sense a stochastic function of z^t, because of functional form of its dependence on the agents' moves we can use Monte Carlo-like techniques to estimate various aspects of $R(z^t)$. In particular, we can estimate its gradient this way, and then have the dynamics use that information to increase R's value from one timestep to the next, hopefully

[60] By our choice of R, here we are only considering distributions over M that have all L of the variables statistically independent. Doing so exponentially reduces the dimension of the space over which we perform the gradient ascent, compared to allowing arbitrary distributions over M. However there may be other restrictions on the allowed distribution that results in even better performance. In the translation of the gradient ascent of $R(z)$ into a collective discussed below, such alternative stochastic forms of the distribution over m would correspond to having agents each of whose moves concerns more than one of the m_i at once.

[61] To establish the first claim, simply note that z^* is a delta function, To establish the second, note that the gradient component $E(R \mid m_i = j)$ is just the expected value of R under a different distribution, z', where z' and z are equal for all components not involving M^i, but z' has a delta function for those components. Since expected R under *any* distribution is bounded above by $R(z^*)$, it must be for z'. Accordingly, each of the components of the gradient is bounded above by $h(m)$, which establishes the claim.

[62] Strictly speaking, we need to encode in either r_i^t or x_i^t the other information specifying the full history, e.g., the values of $z^{t'}$ for $t' < t$. Otherwise that pair of coordinates do not form a complement pair. For completeness, we can choose to encapsulate all such information in r_i^t, as the current value of the seed of an invertible random-number generator used for the stochastic sampling that drives the dynamics (see below). None of the analysis presented here depends on this choice though.

reaching the maximum by time T (in which case we have ensured that G is maximized).

More precisely, at the end of each step t, each agent (i, t) independently samples its distribution z_i^t to choose one of its actions $m_i \in M^i$. That set of L samples gives us a full vector m^t. Next, we evaluate a function of m^t, indexed by (i, t), whose expectation (according to z^t) is the private utility for that agent. (Note that the joint-action m^t is *not* the joint-move of the agents at time t. That is z^t.)

Combining that function's value with other information (e.g., the similar values for i for some times $t' < t$) provides us a training set for that agent controlling variable i. This training set constitutes the worldview for agent $(i, t+1)$, $n_i^{t+1} \in \nu_i^{t+1}$, and is used by the learning algorithm of agent $(i, t+1)$ to form a new z_i^{t+1}. This is done by all L agents, giving us a z^{t+1}, and the process repeats.[63]

This dynamics produces a sequence of points $\{m^t\}$ in concert with a sequence of distributions $\{z^t\}$, which (if we properly choose the private utilities, learning algorithms used to update the z_i, etc.) will settle to m^* and $\delta(m - m^*)$, respectively. As an example, for all i have the function evaluated at m^t be $h(m^t)$, so that the private utility of each agent (i, t) is $R(z^t)$. Have the associated training set for (i, t) be a set of averages of $h(m)$, one average for each of the possible m_i. Have the average for choice $j \in M^i$ be formed by summing the previously recorded $h(m)$ values that accompanied each instance where m_i equalled j, where the sum is typically weighted to reflect how long ago each of those values was recorded. So each of the $|M^i|$ components of n_i^t is nothing other than a (pseudo) Monte Carlo estimate of the components for variable M^i of the gradient of $R(z)$ at the beginning of timestep t.[64] In other words, they are estimates of the components of the gradient of the private utility at the current joint-move.

Accordingly, let the learning algorithm for each agent $(i, t+1)$ be the following update rule:

$$z_i^{t+1} = z_i^t + \alpha \left[n_i^{t+1} - \frac{\sum_j n_{i,j}^{t+1}}{|M^i|}(1, 1, \ldots) \right],$$

where the term in square brackets is the projection of $\nu_{i,t}$ onto its unit simplex $S^{|M^i|}$, the vector $(1, 1, \ldots)$ being normal to that simplex. To keep z in its unit simplex, have α shrink the shorter the distance along $\nu_{i,t}$ from z_i^t to the edge of that associated simplex, $S^{|M^i|}$. The result is that each variable in the collective performs a Monte Carlo version of gradient ascent on G and therefore on h. Moreover, the learning algorithm is a reasonable choice for an agent i trying to

[63] A faster version of this process has all of the agents at a given time share the same m rather than each use a new sample of z^t. This can introduce extra correlations between the moves of the agents though, which may violate our assumption of statistical independence among the $\{M^i\}$.

[64] It would be exactly Monte Carlo if not for the steps updating the $\{z^{t' < t}\}$. It is to account for that updating that the data going into the training set is aged.

modify its move z_i^t to increase its private utility. Accordingly, we would expect it to obey the first premise.[65]

Note that maximizing G is just a problem in design of collectives. This suggests many modifications of the scheme outlined above. In particular, one might try many other learning algorithms besides Monte Carlo gradient ascent to try to find the z that maximizes G. For example, in a **Boltzmann learning algorithm**, each z_i^t is given by a Gibbs distribution over the $|M^i|$ possible values of its variable, with the $|M^i|$ "energies" going into that distribution given by the components of ν_i^t. Using the sampling scheme with this distribution may be better than gradient ascent if the tendency of the latter to get trapped circling local maxima is a concern (say due to the inaccuracy inherent in the Monte Carlo estimating of that gradient). Similarly, one can use many private utilities besides R, in particular ones that try to exploit the first premise. Moreover, all such approaches can be used even if the G and the z's are not an expected utility and associated probabilities over categorical spaces, respectively. The idea of inserting learning agents into a search problem to recast it as a problem in the design of collectives is much more general.

As an example, return to the gradient ascent learning algorithm, and consider replacing $h(m)$ with some $h^*(m)$ that is factored with respect to h for variable i. This will result in a new R, R^*. The partial derivatives of R^* with respect to the $|M^i|$ components associated with the value of variable i equal the corresponding derivatives of R, up to an overall additive term that is independent of m_i. Accordingly, if we set z_i to maximize R^* rather than R, while having all other coordinates still maximize R, we will arrive at the exact same optimizing distribution over m.

Extending this, we can have each coordinate use an associated R^* based on an h^* that is factored for that coordinate, and it will still be the case that if each z_i is set to maximize the associated R^* we end up with the same delta function over m as if all coordinates were set to maximize R. However there is one crucial way that use of R^*'s differs uniform use of R. This arises from the fact that rather than ascending the exact gradient, we are ascending a Monte Carlo estimate of it. That estimation necessarily introduces noise into the ascent. If we can minimize that noise, the ascent should be much quicker. This in fact is exactly what is done when we chose the h^*'s to each have as small ambiguity as possible.[66]

[65] Note that the updates are invariant with respect to translations upward or downward of the function h, since such a translation of h induces an identical translation in R and therefore in n_i^t. Similarly, so long as there are at least two j for which the associated $n_{i,j}^{t+1}$ have different values, $z_i^{t+1} \neq z_i^t$; the updating never halts. This reflects the fact that there are no local maxima.

[66] There are other ways of affecting ambiguity besides the choice of private utility of course, and they have to be traded off other factors in general. As an example, optimizing the step sizes of the agents depends on associated ambiguities. If the stepsizes used by agents other than i are too big, then the gradient estimate for coordinate i will be a poor approximation to the true direction of maximal ascent. To see this, note that if the stepsizes used by agents other than i are too big, then the actual context r for agent i at timestep $t + 1$ will differ significantly from the r at the timestep t. However it is that latter r that determines the value

From this perspective, the idea of casting a search problem as a problem in design of collectives can be motivated as a way to extend gradient ascent so it can be used with categorical variables, by transforming the search to be over a numeric space. Furthermore, even if the underlying space is numeric, casting the search problem as a problem in design of collectives has the advantage over gradient ascent that it naturally allows for large jumps in that underlying space, whether the original space is categorical or numeric, the recasting has the advantage that it allows the search to be decomposed, into a set of parallel searches (one for each agent). If desired, those parallel search can then be implemented on a parallel computer.

More generally, there is nothing about this decomposition that restricts its use to cases where the original global search algorithm is gradient ascent. So in particular, the decomposition can be used directly over a categorical space, without first transforming the search to a numeric space. Moreover, the search/learning algorithms of the individual agents in the decomposition need not be direct analogues of the original global search procedure. So in particular, those individual algorithms need not restrict their agents to only change their states by an infinitesimally small amount, as in gradient ascent. All of these extra capabilities flow from recasting the search problem as a design of collectives problem.

Another modification of vanilla gradient ascent dynamics follows from noticing we are only estimating the gradient of R, rather than evaluating it exactly, and that the estimation is a variant of Monte Carlo. These observations make it natural to modify gradient ascent dynamics by inserting a simulated-annealing-style keep/reject procedure at the end of every timestep. However we cannot do the naive thing, and run that keep/reject procedure on the pair of (the value of $R(z_t)$ before timestep t's modification to z_t), and (that value of R after the modification). This is because we can no more evaluate R exactly than we can its gradient. However we *do* know what the value of h is for the starting m of timestep t and for the new m generated in that timestep. So we can run the keep/reject procedure based on those two values of h.

In fact, we can we can always insert such a simulated-annealing-style keep/reject procedure at the end of each timestep, regardless of the private utility function and/or learning algorithm. This is exactly what is done in the technique of Intelligent Coordinates (IC), sometimes called "Computational Corporations" [26]. From the perspective of design of collectives, IC was motivated as a way to improve techniques that focus exclusively on terms 2 and 3 in the central equation (e.g., by the setting of the private utility). By its insertion of a keep/reject procedure, IC boosts the performance of such techniques by leveraging term 1 in

n at timestep $t + 1$. So having those stepsizes too large means that $P(r \mid n)$ will be broad. This in turn usually induces broad distributions over agent i's private utility values for each of its candidate moves. Usually this means that the ambiguity is quite large.

Conversely, if the stepsize of agent i is too small, then it will be slow in increasing the value of its private utility. So while agent i benefits from having the stepsizes of other agents be as small as possible, its stepsize cannot be too small. Since this holds for all agents, we have to trade off the two effects when determining the optimal stepsize.

the central equation while not degrading terms 2 or 3. Another way of viewing IC is as a variant of a conventional simulated-annealing-style keep/reject search algorithm. In this variation each searched variable is made "smart", its exploration values being the moves of game-playing computer algorithms (agents), rather than as in conventional algorithms, to random samples of a probability distribution.[67]

As a final example of an approach to optimization suggested by extending this gradient ascent example, consider replacing the gradient term with the move of a learning agent in the gradient update rule, rather than replacing the z_i^t term. There are several subtleties with implementing such an idea in practice [9]. One is that typically the value of a utility will change with t even if all the agents freeze their moves with this new approach, since such freezing means that the agents are traversing the surface, only in a constant direction. This contrasts with the typical case where the learning agents set the $\{z_i^t\}$ directly, and can often result in large ambiguities. Nonetheless, especially when in constrained optimization problems like graph traversal, this alternative might be the approach of choice. (See App. D.)

F General situation where the second premise holds

We will illustrate a case where $\int \mathrm{d}n P(n \mid r, s) P^{[\nu]}(x \mid n) = \int \mathrm{d}n P(n \mid r, s) P^{[\nu,\sigma]}(x \mid n, s)$, and therefore the second premise holds.

Consider the integral $\int \mathrm{d}n P(n \mid r, s) P^{[\nu]}(x \mid n)$ arising in the second premise. Expand the distribution in terms of H, and for simplicity say that H does not depend on n directly. Next suppose that $P(n \mid r, s)$ is relatively peaked for fixed r and s. This provides a scale length of the ambiguity arguments of H, given by how much they vary as n moves across that peak. Say that H is a slowly varying function of its arguments on that scale length. (This is particularly reasonable if ambiguities vary little as one traverses the peak in $P(n \mid r, s)$.) Under these circumstances we can pull the integral over n inside the H to operate directly on the vector of H's n-dependent arguments, i.e., replace

$$\int \mathrm{d}n P(n \mid r, s) \mid n) H_{\{A(\underline{\gamma}_\rho; n, x^i, x^j)\}}(x) \;\; \rightarrow \;\; H_{\{\int \mathrm{d}n P(n|r,s) A(\underline{\gamma}_\rho; n, x^i, x^j)\}}(x).$$

Next, consider each term $\int \mathrm{d}n P(n \mid r, s) A(\underline{\gamma}_\rho; n, x^i, x^j)$ appearing inside the H. If we expand that ambiguity and pull in the integral over n, we get

[67] An analogue of IC is a well-run human corporation, with G the corporation's profit, the players i identified with the employees, and the associated $\underline{g}_{it}$ given by the employees' compensation at time t. The corporation is factored if each employee's compensation directly reflects its effect on G. If each compensation package also has good ambiguity, the employees can readily discern how their behavior affects their compensation. Finally, the exploration/exploitation process is analogous to management's deciding whether to maintain or abandon a particular set of decisions by the employees. These similarities are the basis of the name "computational corporation".

expressions of the form $\int \mathrm{d}nP(n \mid r,s)P(\underline{g}_{\rho,\sigma}(x^1,\rho))P(\underline{g}_{\rho,\sigma}(x^2,\rho))$. Now again assume $P(n \mid r,s)$ is relatively peaked, this time on the scale of variations in $P(\underline{g}_{\rho,\sigma}(x,\rho))$. This allows us to replace

$$\int \mathrm{d}nP(n \mid r,s)P(\underline{g}_{\rho,\sigma}(x^1,\rho))P(\underline{g}_{\rho,\sigma}(x^2,\rho)) \;\rightarrow$$
$$\int \mathrm{d}nP(n \mid r,s)P(\underline{g}_{\rho,\sigma}(x^1,\rho) \mid n)\int \mathrm{d}n'P(n' \mid r,s)P(\underline{g}_{\rho,\sigma}(x^2,\rho) \mid n').$$

Expand the first integral in this product as

$$\int \mathrm{d}r'[\int \mathrm{d}n\mathrm{d}s'_\rho\delta(\underline{g}_{s'_\rho}(x,r') - y)P(r',s'_\rho \mid n)P(n \mid r,s)]$$

(and similarly for the second).

Say that the first distribution in the integrand is peaked, in s'_ρ, about some $h(n)$, and that the second one is peaked about the n lying in the preimage $h^{-1}(s)$. (This is exactly true if n specifies s precisely.) Then we can replace

$$\int \mathrm{d}n\mathrm{d}s'_\rho\delta(\underline{g}_{s'_\rho}(x,r') - y)P(r',s'_\rho \mid n)P(n \mid r,s) \;\rightarrow$$
$$\int \mathrm{d}n\delta(\underline{g}_s(x,r') - y)P(r',s'_\rho \mid n)P(n \mid r,s)$$

We would have arrived at the exact same expression if we had made the analogous approximations in expanding $\int \mathrm{d}nP(n \mid r,s)P^{[\nu,\sigma]}(x \mid n,s)$ instead. Hence these approximations justify the second premise. However the second premise can hold even if not all of those approximations of peaked distributions are valid, so long as there is sufficient cancellation among the contributions from the wings of the distributions (e.g., it will hold if $\nu \subseteq \sigma$ regardless of such peakedness). So the second premise is weaker than these approximations. In fact, under those approximations, we could always replace the ambiguities arising in H with their averages according to $P(n \mid r,s)$, something which we do not do in the current analysis.

G An alternative definition of ambiguity

Note that rather than $P(y^1,y^2;\psi;l,x^1,x^2)$, the difference of the distributions of utility values at x^1 and x^2, one could consider the distribution of differences,

$$P^*(y^1,y^2;\psi;l,x^1,x^2) \equiv \int \mathrm{d}r\mathrm{d}sP(r,s \mid l)\delta(y^1 - \psi_s(x^1,r))\delta(y^2 - \psi_s(x^2,r)),$$

and the associated ambiguity A^*. Now almost all of the theorems and corollaries presented below hold for ambiguities based on A^* as well as A, so we could use A^* rather than A if we wanted to. Moreover, P is P^* modified to preserve the marginals of the random variables ψ^1 and ψ^2 while making those variables be independent:

$$P(y^1, y^2; \psi; l, x^1, x^2) = P^*(y^1; \psi; l, x^1, x^2) P^*(y^2; \psi; l, x^1, x^2).$$

So A^* fixes (P^* which fixes P which fixes) A, but not vice-versa, i.e., A contains less information than A^*. Furthermore, of all ambiguities based on a distribution with the same marginals as P^*, A is the "widest", having the largest region in which it is neither 0 nor 1.

However all of this does not mean that we are just being more conservative by using A rather than A^*, i.e., that we are discarding certain predictions concerning orderings of CDF's that we would make if we used A^*, while keeping other such predictions. That's because in general A can shrink in going from one l to another (i.e., its value can decrease for at least one y and not increase for any y) while A^* does not, and vice-versa.[68] So either choice of ambiguity may result in predictions that would not have been made with the other choice.

In this paper we restrict attention to learning algorithms whose behavior depends on increasing/decreasing ambiguities based on A rather than on A^*. This seems to be the case for most real-world learning algorithms, and therefore A rather than A^* seems to be the appropriate quantity to plug into our results. Only if the learning algorithm exploits information in n about the relation of utility values *at the same* r would changes in A^* be a better predictor of associated changes in what move the algorithm is likely to make. This is rarely the case though. For example, training sets formed in the course of multi-step games (see App. D) contain information about utility values for move/context pairs (one such pair for each preceding timestep), rather than for multiple moves in a particular context.

Despite this though, since A^* fixes A but not vice-versa, parameterizing H in terms of A^* rather than A would make H more flexible. However since the premises only involve A, not A^*, to simply the exposition here we will write H in terms of A.

References

[1] N. I. Al-Najjar and R. Smorodinsky. Large nonanonymous repeated games. *Game and Economic Behavior*, 37(26-39), 2001.

[2] R.J. Aumann and S. Hart. *Handbook of Game Theory with Economic Applications, Volumes I and II.* North-Holland Press, 1992.

[3] T. Basar and G.J. Olsder. *Dynamic Noncooperative Game Theory.* Siam, Philadelphia, PA, 1999. Second Edition.

[68] However it is not possible that A can shrink while A^* increases, since if A shrink that means the difference in the expected values of ψ at x^1 and x^2 decreases while if A^* grows that difference must increase.

[4] R. H. Crites and A. G. Barto. Improving elevator performance using reinforcement learning. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems - 8*, pages 1017–1023. MIT Press, 1996.

[5] J. Eatwell, Milgate M., and Newman P. *The New Palgrave Game Theory.* Macmillan Press, 1989.

[6] Y. M. Ermoliev and S. D. Flam. Learning in potential games. Technical Report IR-97-022, International Institute for Applied Systems Analysis, June 1997.

[7] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge, MA, 1991.

[8] V. Krishna and P. Motty. Efficient mechanism design. (pre-print), 1997.

[9] J. Lawson and D. Wolpert. The design of collectives of agents to control non-markovian systems. In *of American Association of Artificial Intelligence Conference 2002*, 2002.

[10] R. D. Luce and H. Raiffa. *Games and Decisions.* Dover Press, 1985.

[11] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theorey.* Oxford University Press, New York, 1995.

[12] D. Monderer and L. S. Sharpley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.

[13] N. Nisan and A. Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35:166–196, 2001.

[14] M. Osborne and A. Rubenstein. *A Course in Game Theory.* MIT Press, Cambridge, MA, 1994.

[15] D. C. Parkes. *Iterative Combinatorial Auctions: Theory and Practice.* PhD thesis, University of Pennsylvania, 2001.

[16] P. Tucker and F. Berman. On market mechanisms as a sofware techniques. Technical Report CS96–513, University of California, San Diego, December 1996.

[17] K. Tumer and D. H. Wolpert. Overview of collective intelligence. In D. H. Wolpert and K. Tumer, editors, *The Design and Analysis of Collectives.* Springer-Verlag, New York, 2002.

[18] D. H. Wolpert. The lack of a prior distinctions between learning algorithms and the existence of a priori distinctions between learning algorithms. *Neural Computation*, 8:1341–1390,1391–1421, 1996.

[19] D. H. Wolpert. A mathematics of bounded rationality. (in preparation), 1999.

[20] D. H. Wolpert and J. Lawson. Designing agent collectives for systems with markovian dynamics. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, Bologna, Italy, July 2002.

[21] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997. Best Paper Award.

[22] D. H. Wolpert and M. Millonas. Experimental tests of the theory of collectives. Available at http://ic.arc.nasa.gov/ dhw, 2003.

[23] D. H. Wolpert and K. Tumer. Optimal payoff functions for members of collectives. *Advances in Complex Systems*, 4(2/3):265–279, 2001.

[24] D. H. Wolpert and K. Tumer. Collective intelligence, data routing and braess' paradox. *Journal of Artificial Intelligence Research*, 2002. to appear.

[25] D. H. Wolpert, K. Tumer, and J. Frank. Using collective intelligence to route internet traffic. In *Advances in Neural Information Processing Systems - 11*, pages 952–958. MIT Press, 1999.

[26] D.H. Wolpert, K. Tumer, and E. Bandari. Intelligent coordinates for search. 2002. submitted.

[27] G. Zlotkin and J. S. Rosenschein. Coalition, cryptography, and stability: Mechanisms for coalition formation in task oriented domains. (pre-print), 1999.