# Games and Kinds

Cailin O'Connor


## Abstract

In response to those who argue for 'property cluster' views of natural kinds, I use evolutionary models of sim-max games to assess the claim that linguistic terms will appropriately track sets of objects that cluster in property spaces. As I show, there are two sorts of ways this can fail to happen. First, evolved terms that do respect property structure in some senses can be conventional nonetheless. Second, and more crucially, because the function of linguistic terms is to facilitate successful action in the world, when such success is based on something other than property clusters, we should not expect our terms to track those clusters. The models help make this second point salient by highlighting a dubious assumption underlying some versions of the cluster kinds view—that property clusters lead to successful generalization and induction in a straightforward way. As I point out, those who support property cluster kinds as natural can revert to a promiscuous realism in response to these observations.

**Contents**



## 1 Introduction

The ‘property cluster’ view of kinds provides an alternative to accounts of natural kinds on which necessary and sufficient conditions must be given for kind membership. On the cluster kinds picture, groups of objects in the world that have similar properties as a result of homeostatic mechanisms, or historical processes, may be proper natural kinds even if they cannot be defined using necessary and sufficient conditions. In many cases, because the clustering of properties leads to certain successes of induction or generalization in treating the sets of objects as a kind, we expect our linguistic terms to track these clusters (Millikan to appear; Boyd 1991). In this way the view alleviates worries that kind language is merely conventional. While it may be difficult to give conditions for kind membership, we can still trust that our kind terms, in many cases, track something real.

This paper will use tools from evolutionary game theory to assess the claim that evolved terms track such cluster kinds. In particular, I employ the sim-max game, first introduced by Jäger (2007), to model language acquisition in the sort of world described by proponents of the cluster kinds view. On first glance, previous results from these models seem to strongly support the cluster kinds position. Terms evolve that seem to group similar objects, and that seem to respect property clustering. A closer look at the models, though, muddies this picture. First, evolved terms that do respect property structure in some senses can be conventional nonetheless. Second, and more crucially, because the function of linguistic terms is to facilitate successful action in the world, when such success is based on something other than property clusters, we should not expect our terms to track those clusters. The models help make this second point salient by highlighting a dubious assumption underlying some versions of the cluster kinds view—that property clusters lead to successful generalization and induction in a straightforward way.

There is a response for the cluster kinds theorist. As I point out, there is not actually some canonical way to pick out properties that are the objectively natural ones. The properties we care about are usually the ones that are already relevant to our payoff in some way. Furthermore, terms will generally be natural with respect to some properties—those that matter for the relevant inductive success—and not with respect to others. So the cluster kinds theorist can say simply that cluster kinds are natural kinds with respect to just some set of inductively relevant properties. This yields a picture of the naturalness of cluster kinds related to John Dupre's promiscuous realism. To be clear, I do not necessarily endorse this view, but merely point out that it is available. On the other hand, one might also point out that for many terms, both in the sciences and everyday life, the properties they will be natural with respect to are ones we usually think of as highly unnatural, such as 'ability to stimulate sweetness receptors' or 'ability to cause vomiting in humans'.

The real take-away from the analysis of sim-max games here, then, is a deeper understanding of where and when the cluster kinds picture will be a useful one—in those cases where the properties we are concerned with are relevant to the inductive success of the kind at hand.

The paper will proceed as follows. In section 2, I introduce two prominent accounts of cluster kinds, the first from Richard Boyd and the second from Ruth Millikan. In section 3, I describe the basic models employed here and relevant results on the evolution of terms in these models. In section 4, I address in detail two sorts of ways that kind terms might fail to respect property clustering. In section 5, I argue that appealing to the role of perception in determining linguistic term application does not help the problem. And in section 6 I discuss a response for cluster kinds theorists to the worries that sim-max games raise.

## 2 Cluster Kinds

Let us start with an oft employed framework, or intuition pump, for understanding property cluster kinds. Imagine a multidimensional space where each dimension corresponds to a property that real-world objects might possess. There are dimensions for shape, size, surface reflectance, hardness, surface texture, scent (or rather the sorts of molecules the object puts off), chemical composition, and so on. Using this space, any object at a particular moment in time can be situated within it.[1] Later in the paper, I will discuss this construct in more depth, but for now let us take it at face value.

One might also imagine a space with dimensions for 'social properties', like aggressiveness or dominance or level of liberty for a country, where social kinds like 'alpha male' or 'democracy' could be situated. Boyd (1991, 1999) defends such social kinds as appropriate candidates for natural kindhood. I will not worry about these cases for now.

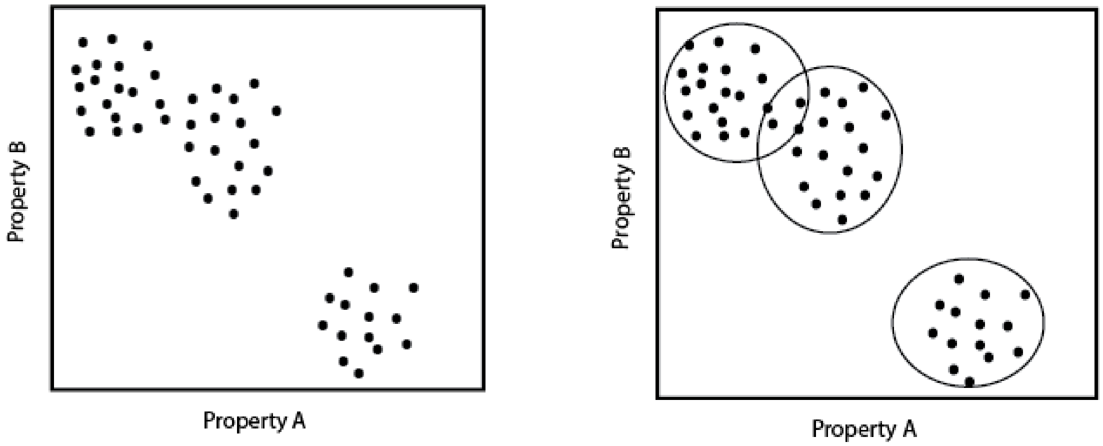
Cluster kinds will be sets of objects that tend to cluster together within this space as the result of some process. On the picture presented by Boyd (1990) clustering of properties is a result of, 'a sort of homeostasis. Either the presence of some of the properties...tends (under appropriate conditions) to favor the presence of the others, or there are underlying

[1] Over time, of course, most objects will wander in the space. A peach will shift on many dimensions as it grows, matures, and eventually decays.

mechanisms of processes that tend to maintain the presence of the properties...or both' (373). Millikan points out that, 'Most of the physical objects in which we take an interest have arrived in our world through some process of reproduction or, in the modern world, mass production' (Millikan to appear, ch.1). The implication is that reproduction and mass production are historical processes that tend to generate objects with similar properties. Manmade objects are often created using a template and so tend to resemble manmade objects of the past. Biological objects are produced through reproduction and tend to resemble their forebears.[2] For both authors, these clusters need not be perfectly separable—there may be overlap with other clusters, or vagueness at their boundaries. They also need not stay in the same area of property space—clusters can migrate, as biological species do over the course of evolution.

Figure 1 (a) shows a sample visual representation of the sort multidimensional space described. This representation only contains two dimensions so that it can be easily displayed here. In this representation we see a space with three property clusters. Two of these have some overlap. Perhaps these are blackberries and black raspberries, while the other cluster is the set of pears.

[2] Other objects in the natural world (think rocks, mountains, rivers) clump together as a result of immutable physical laws.




(a) (b)

Figure 1: A representation of clusters of objects in property space (a), and terms that might cover these objects (b). Dots represent objects and circles the terms that cover them.

What is the actual claim about linguistic terms that this space is used to motivate? It is something like this: when real-world objects form clusters of this sort, as a result of a historical process, it will often be the case that treating them as a kind leads to success in induction and generalization. If we treat blackberries as a kind, for the purposes of eating, or for the purposes of science, we will be able to generate inductive success. Eating the blackberry worked today. If we suppose that we can always eat them, we will be successful. In other words, we should think our terms attach to categories of things that may resist definition using necessary and sufficient conditions, but are nonetheless unified as property clusters, and thus are appropriate categories for induction and generalization. Figure 1 (b) shows a representation of possible terms that might cover these objects, with each circle representing the area of property space that a linguistic term covers. In this case, we see that two terms have some space of overlap, preventing necessary and sufficient conditions for membership in these categories, but they nonetheless seem intuitively natural.

In this short description, I have abstracted away from differences in the views of the two prominent cluster kinds theorists discussed. Boyd (1991) gives an early defense of the cluster kinds view. At the heart of Boyd's argument is the appeal to inductive success described above. He argues that property cluster kinds are defined as a result of inductive success in the sciences based on treating the kind as such, that this success is due to some underlying structure of the world related to the cluster of properties, and that scientific terms refer to these clusters as a consequence of this inductive success (Boyd 1989, 1990, 1991). The paradigm example of these kinds, which he sometimes refers to as homeostatic property cluster kinds, is the species, where biological facts related to reproduction and development lead individuals in a species to have clusters of properties that are relevant for the inductive success of science, despite natural variation and speciation often preventing the identification of necessary and sufficient conditions for species membership.[3]

Millikan (to appear) gives a much broader analysis of the ontology of property cluster kinds

[3] To be perfectly clear, in some papers Boyd seems to think that use of a successful kind term is evidence that members of the kind do, in fact, have clusters of relevant properties, in other papers he defines cluster kind terms as just those successful inductive categories that happen to correspond to actual clusters of properties (leaving open the possibility that some do not). Boyd (1999) seems simply to define natural cluster kinds as whatever corresponds to a term which generates inductive success, regardless of the properties of the referents. From later descriptions of his position in this paper (for example, see, Boyd (2010)) I gather that he in fact thinks that inductive success must be linked with some sort of cluster of properties. Going forward, I will assume that Boyd has a picture of cluster kinds where their inductive success is based in shared properties.

and their relationship to language. Inductive success is central to her account, as it is for Boyd, but she thinks that property cluster kinds just are the real world structure that accounts for the inductive success of organisms (whereas Boyd considers them as one sort of kind that is relevant to success in the sciences in particular). Millikan, like Boyd, limits her kinds to property clusters where the clustering occurs for some ‘univocal reason’ such as a historical process. For example, weasels on twin Earth, even if they are identical to weasels produced on Earth, will not qualify as members of the same kind. As a general point, these two authors illustrate that property cluster kinds can be taken to be a broader or a narrower phenomenon. Millikan treats cluster kinds as much more ubiquitous and more crucial to cognition and language than Boyd seems to.

Now we have a basic picture of the theory of property cluster kinds. In the next sections I will ask the following question: using evolutionary models, what can we say about whether and when terms will, in fact, track such property clusters?

## 3 Models and Results

How does one go about assessing whether or not linguistic terms do, in fact, track objects that share properties? One method is to understand generally, given the functional role that language plays in human communities, the conditions under which this should be expected to occur. This approach will be employed here. In particular, I use evolutionary game theory to model the evolution of linguistic categories, and then employ these models to discuss the issues at hand.

Why, one might ask, is this modeling framework useful or appropriate in this case? Linguistic terms develop in a social context. In other words, they generally arise as a result of interaction between humans who are attempting to communicate. This communication serves a functional role in the lives of the humans employing it. It allows humans to transfer information about the world that then guides their future action. In other words, human communication is strategic in the sense that it involves multiple actors and at least some of these actors really care about the success of this communication. Furthermore, language arises through processes of cultural evolution and individual learning. Evolutionary game theoretic models embody both of these features. First, they focus on human interaction that has a strategic component. And second, they model behaviors that arise as a result of an evolutionary process, including processes of cultural evolution and learning.

### *3.1 Sim-Max Games*

Jäger (2007) introduced the sim-max (similarity-maximizing) game to model the evolution of linguistic categories. In particular, he uses this game to defend claims by Gärdenfors (2000) that concepts will generally attach to convex areas of perceptual space. (More on this later.) Since this introduction, Jäger *et al.* (2011) expanded formal work on the sim-max game, and it has been used by a number of philosophers to model the evolution of language (Franke *et al.* 2011; O'Connor 2013, 2015a; Franke and Correia to appear) and perceptual categorization

(O'Connor 2014).

The sim-max game is a version of the signaling game introduced by Lewis (1969). The Lewis signaling game involves two actors—a sender and a receiver. It proceeds as follows. First the sender observes the state of the world. She then sends a signal to the receiver contingent on this observation. The receiver is then able to take an action contingent on receipt of this signal. The goal of both actors is to coordinate the act taken by the receiver with the state observed by the sender. If this occurs, they both receive a payoff. In other words, different states of the world require different actions, and the signal in this game is a communicative tool the actors can use to get the right match. To give a brief example, the actors might be a daughter who regularly drives to take her elderly mother on a walk. The states might be that it is either sunny or raining at the mother's house, and the actions might be for the daughter to either pack sunscreen or umbrellas. If the two actors can use signals, in this case a phone call, to coordinate sunscreen with sunny days and umbrellas with rainy ones, they both get a payoff. The sim-max game adds extra structure to this base framework. In particular, the states of a sim-max game exist in a space where distance in the space represents similarity. For example, the state space of a sim-max game might be five states arranged on a line. Because state one is closer to state two than to state five, we know that states one and two are more similar to each other than states one and five. Modifying the above example to fit this case, the possible states of the world might be a downpour, a heavy rain, a light rain, a cloudy day, or a sunny one. We can see here how the state space now incorporates similarity—if we don't think too carefully about ‘similarity’, we can say that a downpour is more similar to a heavy rain than it is to a sunny day.[4] Figure 2 shows a representation of this particular state space. Sim-max games can have state spaces of any dimensionality, and thus have the potential to represent the sort of situation cluster kinds theorists are interested in. Objects exist in a multidimensional space that represents the similarity of objects, and humans learn to attach linguistic terms to these (clusters of) objects.

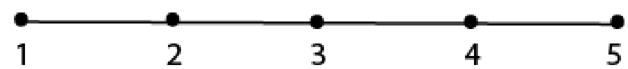



Figure 2: A state space for a sim-max game with five states arrayed on a line.

In the sim-max game, similarity is hashed out in terms of payoff. As in the basic signaling game, each state is associated with some ideal action. These actions also work well, though, when taken in similar states of the world. For example, the action that works for the downpour (state one) in figure 2—to bring umbrellas, heavy raincoats, and galoshes—will also work decently well in state two because appropriate actions for a downpour will also be appropriate for a heavy rain. This action will work less well in state three, where the ideal might be to bring just an umbrella and not that other heavy, hot stuff. The action will be worse in four, and worst of all in state five. An assumption that will be employed for all the games discussed

[4]Or we can follow Goodman (1972) and think of similarity as being with respect to some chosen property, in this case the amount of water falling from the sky.

here is that the further two states are from each other in the state space of a sim-max game, the greater the loss in payoff for taking the action appropriate to one of them in the other state.[5] Note that actors in a sim-max game, unlike those in the Lewis signaling game, typically have a double task. First, they must break states into categories for the purposes of signaling and second they must conventionally attach signals to each category. In our weather example, there are enough linguistic categories to describe all five states. But suppose we instead used a game to represent this situation that had 100 states arrayed on the same line, or 1000, where each state corresponded to a slightly different level of rain. In such a case, in order to be successful, our actors would have to simultaneously divide these states into relevant categories, and attach terms to them. States 0-10 might count as a 'downpour', 10-30 as 'rainy', etc.

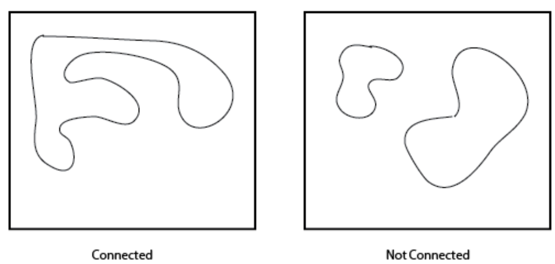
In the next section, I will discuss basic results on the evolution of term language in sim-max games. But one question that will be useful to answer before continuing is the following. What do 'natural' terms look like in these spaces? When asking whether term evolution respects natural structure, what are we even asking?

There are a few desiderata that one might identify for natural kinds terms modeled in this way. The first, and perhaps most important, is picked out by Gärdenfors (2000) in related work.[6]

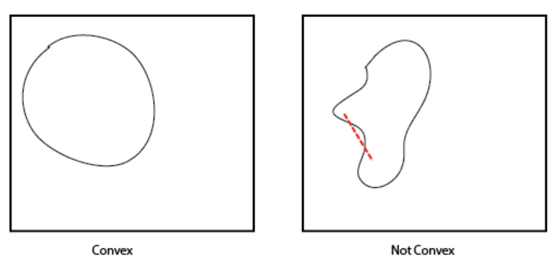
[5]In many cases, modelers use a function to model this payoff loss. The distance between the two states is calculated and then a function determines payoff to both actors. This function can take many forms (quadratic, gaussian, linear), but has generally been assumed to be strictly decreasing in distance.

[6]In particular, he attempts to say something about which concepts pick out natural areas of conceptual spaces. Later in the paper, I will say more about what conceptual spaces are.

He identifies convex categories as the natural ones in spaces similar to those we are looking at. A convex area is one where for any two points within the area a line between them will remain inside it. Convexity also implies connectedness. Connected areas are those such that for any two points inside the area a path can be drawn between them that stays inside it. Figures 3 shows examples of connected and convex terms, and non-connected and non-convex terms, in a two dimensional state space. The line in (b) does not remain within the figure, demonstrating that the figure is not convex.




(a)

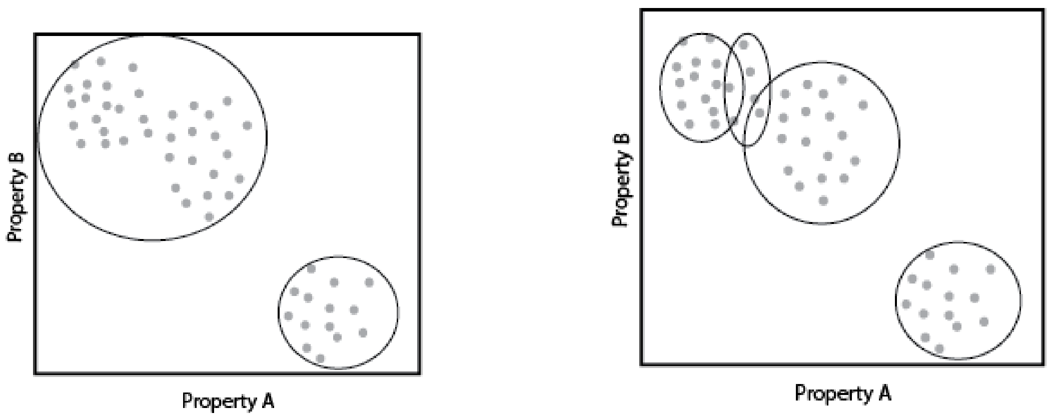



(b)

Figure 3: Examples of connected, non-connected, convex, and non-convex areas of a space.

Another desiderata is that terms respect the sorts of clumps that Millikan argues generally occupy property space. Terms that do this will not group together two clumps, or separate one. Figure 1 (b) gets it about right. Figure 4 (a) shows terms that group two clusters together.

(b) shows terms that separate one cluster into two categories.




(a) (b)

Figure 4: Examples of terms that cover multiple clusters (a) or divide one cluster (b). Circles represent the area covered by a linguistic term.

One last desiderata for kind terms is that they not be conventional. This requirement is often emphasized by natural kinds theorists—kinds must carve the world at its joints. The implication, of course, is that there exist clear joints and a canonical way to carve them. Conventionality, on the other hand, implies some level of arbitrariness—that the division could have been otherwise.

It is worth noting that there is plenty of room for disagreement as to whether and which of these desiderata are actually necessary for terms that pick out natural regions of property space. Probably the most serious offenders are terms that are not connected, followed by those that are not convex. When it comes to property clusters, it may be unclear whether there is even a canonical way to respect clustering, or a non-conventional way to derive category boundaries. Keeping these limitations in mind, in the rest of the paper, I will use these desiderata to ground our discussion of how evolved terms can fail to respect property structure.

### *3.2 Voronoi Languages, Clusters, and Categories*

Sim-max games are the model that will be used to represent term evolution in property cluster spaces, and so to assess whether and when linguistic terms are expected to track clusters. The question is now: what sorts of terms evolve in these games? As we will see, on first glance these terms seem to almost exactly mimic what cluster kinds theory predicts, though a closer look will complicate the picture.

To understand evolution in sim-max games, it will be useful to first say a word about optimality in these models. For simplicity sake, let us start with situations where states are distributed evenly in the state space of the game, and then move on to clusters. Sender and receiver both want to use strategies that, on average, make the action taken by the receiver as good as possible for the state of the world. In the sim-max game, this means that the best they can possibly do is to use a single term for each possible state, so that upon receipt of this term the receiver is completely sure which state of the world obtains and is able to pick the perfect action. In other words, optimal languages do not categorize, but treat each possible state of the world as unique.[7] As I argued with the rain example, though, this will not always be

[7] I should be careful here. Because similarity tracks payoff in sim-max games, any states for which there is no payoff relevant difference in terms of action will be the same state in the model. In other words if two rainy days that differ slightly can be responded to in the exact same way without influencing payoff, they will be the same state in the sim-max game. So

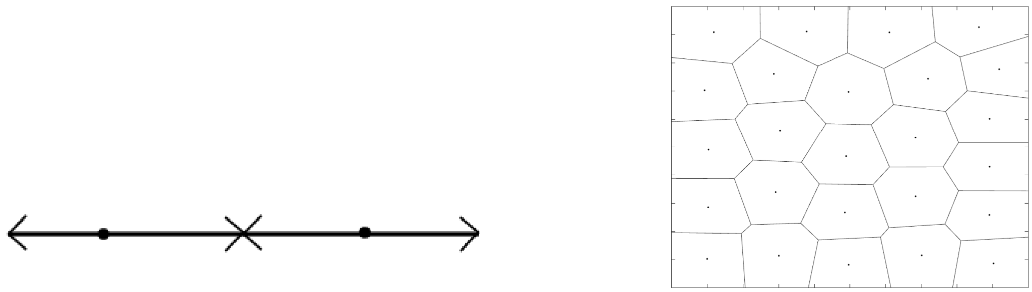
feasible. It is often the case that there are many, many separate states that actors must respond to. In these cases, it will be too cognitively costly to agree on, and keep track of, a separate term for each one.[8] For this reason, analyses of sim-max games usually assume that the number of terms is limited compared to the number of states.

In these cases optimal languages—those that provide the best possible payoffs for the sender and the receiver by minimizing the average distance between the state of the world and the act taken—are what Jäger *et al.* (2011) call Voronoi languages. Voronoi languages divide the state space into convex areas of approximately equal size and assign a term to each area. Upon receipt of a term, the receiver takes the action that is best for a state right in the middle of the area the term covers. Figure 5 shows examples of such a language for a one dimensional state space and a two dimensional state space. Each cell represents the states covered by a single term. The dot in each cell represents the receiver's action for that term.

---

when I claim that each state has a unique term in an optimal language, I mean that each payoff relevant state has a unique term.

[8]O'Connor (2015a) proves that for sim-max games with many states even small cognitive costs will limit the number of terms in an efficient language compared to the number of states.



(a) (b)

Figure 5: Voronoi languages in sim-max games. Figure (a) shows a Voronoi language for a one-dimensional state space with two terms. The state space is divided in half, with each half represented by one term. (b) shows a Voronoi language for a two dimensional state space with 36 terms. Terms pick out roughly equally sized areas of state space.

Why is this sort of language optimal? Let us start by discussing sender optimality. A receiver strategy assigns an action to each signal they might receive in a sim-max game. For example, in the rain example suppose that there are two terms—rainy and sunny. Suppose that given the signal 'rainy' the daughter brings umbrellas and given the signal 'sunny' she brings sunscreen. For this receiver strategy, if the state is such that sunscreen is the best of the two possible actions, the sender should pick the signal 'sunny' for an optimal outcome. In doing so, she makes the receiver's action most appropriate for the state, and so maximizes her payoff.[9] What about receiver optimality given optimal behavior by a sender? The receiver does best to also try to minimize the distance, on average, between the state of the world and the action she takes (and so maximize payoff). She does this by spreading out her actions in

[9]Those familiar with game theory will notice that I only consider pure strategies here. This may seem like a strange choice, but Lipman (2009) shows that vagueness, as represented by mixed strategies, is always suboptimal in common interest signaling games. For the sim-max games we consider here this is true for all except knife's edge cases, and so I ignore mixed strategies.

such a way that the resulting categories are about equally sized.[10] In our example, it would make no sense to use the signal 'rainy' and 'sunny' to prompt actions appropriate to the most extreme cases, say bringing full rain gear in response to 'rainy', and sunscreen, sun hats, and sun dresses in response to 'sunny'. The actors do better if the receiver picks actions that are more average for each category, because these will be closer to appropriate for more of the states in the category. The particulars of how to do this will depend on the state space, but generally the picture is that in a Voronoi language there are a set of terms that partition the state space about equally, each of which the receiver responds to appropriately.

Evolution, of course, is not an optimizing process. So we can now ask: what happens when actors evolve terms in sim-max games? Although there are many relevant details that I abstract away from, Voronoi languages, or languages that approximate them, are expected to evolve in these games.[11] In other words, in groups of actors evolving to categorize states and communicate about these categories, actors will carve states up into approximately convex sets, of approximately equal size, and attach terms to these. This evolution can be either biological, or, importantly for our purposes, cultural, as the models used to generate these results can be interpreted either way.[12]

[10] I am being handwavy here. By 'evenly sized' term coverage, what I really mean is that each term covers an area of the state space that is about equally important from a payoff point of view.

[11] Jäger (2007) shows that the asymptotically stable rest points of the replicator dynamics for sim-max games will be Voronoi tessellations induced by the receiver strategy. Jäger *et al.* (2011) have similar results for games where the state space is continuous.

[12] For more on the interpretation of the replicator dynamics as representing cultural evolution via imitation learning, see Weibull (1997). For more detailed analyses of the evolutionary prop-

As mentioned, the results just discussed on evolution of terms in sim-max games implicitly assume that actors are dealing with states that are about evenly placed throughout the state space. For each state, there will be states that are highly similar, others that are less similar, still others that are less similar, and so forth. In other words, there are no missing gaps in the state space, where states simply do not occur. For the purposes of this paper, of course, we are also interested in models with clusters of states separated at least in part by empty space as in figure 1.

What happens when actors evolve terms in sim-max games with clusters? Again being a bit handwavy, the further apart two such clusters are in state space, the more likely that they will be grouped into separate categories. Why? Distance in the state space, remember, translates into payoffs. The farther apart states are, the worse the ideal action for one is for the other. This distance makes it increasingly likely that the boundaries between categories in a Voronoi language for the game will fall in between the clusters.[13] Figure 6 shows this visually. Gray dots represent objects, cells represent areas of term coverage, and black dots represent receiver actions. In diagram (a), the Voronoi language involves a term that attaches to states in two clusters. In diagram (b), when these clusters are farther apart, the Voronoi language

erties of these models, see Jäger (2007); Jäger *et al.* (2011); O'Connor (2014). Unpublished work by Elliott Wager, described in O'Connor (2014), is also relevant here. In particular, he shows that some non-convex categorizations in sim-max games will be Lyapunov stable under the replicator dynamics.

[13] Proofs from Jäger (2007) showing that only Voronoi languages will be asymptotically stable under the replicator dynamics for sim-max games with a finite state space are directly applicable to games with clustered states. The observations in this paragraph follow from the character of these languages.

separates them.[14] This said, Voronoi languages will not always respect clustering. A language with too few terms may group multiple clusters under one term to achieve optimality. One with many terms may divide clusters.

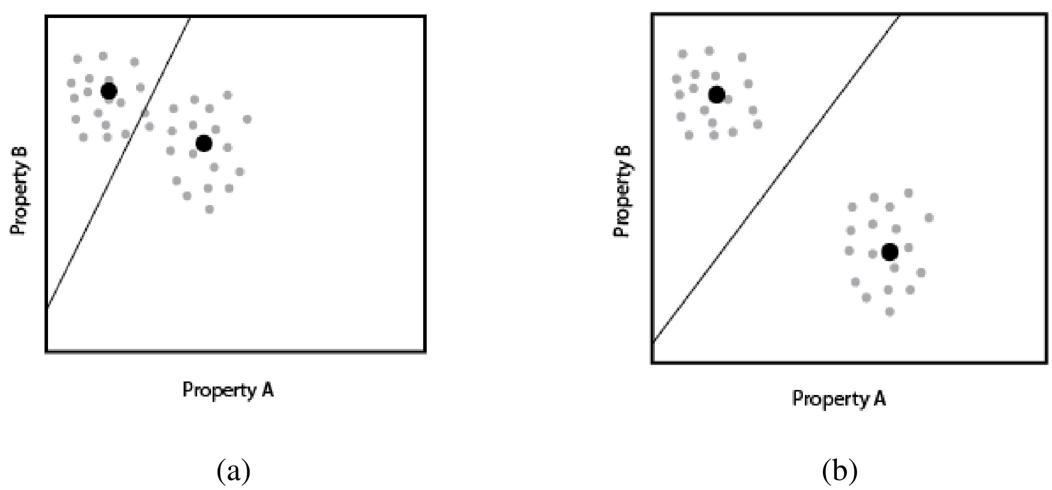



Figure 6: Voronoi languages in sim-max games with clusters of states. (a) shows a language for two close clusters. (b) shows that this language respects clustering if they are further separated in the state space.

There is one last thing to mention before continuing. Real world terms are often vague, meaning that they clearly apply to some states, clearly do not apply to others, and have a boundary region of unclear application. The most paradigmatic cases of vagueness often occur when objects are well modeled using the sim-max game framework—where there are similarity relationships between these objects and a slow transition between some sorts of objects and others, with no gaps in between. Imagine the ‘heap’ in the Sorites paradox. A

[14] In simulations of two actors using Roth-Erev reinforcement learning to form terms in a sim-max game modeled as ten states on a line, I similarly find that the larger a gap in the middle of the state space, the greater the probability that categories respect this gap.

heap with 1000 grains is very similar to one with 999 grains, less similar to one with 900 grains, and still less similar to one with 100 grains. Voronoi languages, though, do not display vagueness. They partition state spaces neatly. Previous work by philosophers has explored the evolution of vague terms in these games, showing that bounded cognition (Franke *et al.* 2011), generalized learning (O'Connor 2013), or confusion about states (Franke and Correia to appear) can lead to vague categories. Importantly, though, these categories still tend to qualitatively look like Voronoi languages, just with fuzzy borders.[15]

To summarize: previous results in sim-max games indicate that terms should evolve to cover convex regions of states, and to generally respect clustering. (With, of course, the caveat that this is an approximate prediction.) At first glance, these results seem to perfectly support claims in cluster kinds theory. In the next two sections, however, I will problematize this view by discussing the role that convention and function play in shaping categories.

## 4 Conventionality and Functionality in Kind Terms

In this section, I will address two ways that evolved terms in sim-max games might fail to 'carve nature at its joints' when it comes to property cluster kinds. The first involves the conventionality of categorization in signaling models. The second, and more important, involves the functionality of categorization in these models.

[15] Franke and Correia (to appear) also find that in cases where vague terms evolve, there is more regularity to the evolution of the language.

### *4.1 Conventional Categories*

Barrett (2007) uses signaling games to argue that natural kind terms may be conventional, rather than tracking some deep real-world structure. The simplest of his models works as follows. There are four different states. Two senders are able to observe these states and each may send a signal contingent on this observation. Let us call the signals for the first sender ‘red’ and ‘blue’ and those for the second sender ‘thick’ and ‘thin’. The receiver gets a pairing of signals from the two senders, say ‘red and thin’, or ‘blue and thin’, and takes an action contingent on the signal pairing. If the action is the correct one for the state, the senders and the receiver all get a payoff. As Barrett shows, when the actors use a very simple learning rule to update their behaviors, they often end up with a system where the first sender’s terms partition the states into two sets, and the second sender’s terms do the same, but in a different way.[16] This allows the receiver to guess the correct state of the world based on these partitions. Figure 7 shows a visualization of two such strategies. Crucially, there are multiple ways for the actors to successfully form these partitions. (a) shows a system where sender one attaches terms to 1-2 and 3-4, while sender two attaches terms to 1-3 and 2-4. (b) shows a system where sender one still groups 1-2 and 3-4, but sender two groups 1-4 and 2-3. In other words, the world does not fully determine successful term application. As Barrett argues, simple actors using these terms might become convinced that they track real-world structure

[16] Barrett looks at a few learning rules in this paper, all based off Roth-Erev reinforcement learning (Roth and Erev 1995).

as a result of their completely successful use, but from a God's-eye-view it is clear that the term categories are arbitrary and so arguably do not carve the world at its joints in a meaningful way.


Figure 7: Two examples of evolved languages in Barrett's natural kinds models. In (a) and (b), sender one partitions the states in the same way, but sender two chooses different partitions.

In sim-max games, category boundaries may, likewise, be conventional in several different ways. Again, let us begin the discussion with games where states are evenly distributed in the state space, then extend these lessons to games with clusters. First, even when actors evolve perfect, optimal Voronoi languages, these can be conventional in that they can divide state spaces along different axes (Jäger *et al.* 2011). Consider figure 8. This shows two Voronoi languages for a two-dimensional state space. Either of these can evolve (Jäger *et al.* 2011). They will be equally good from the point of view of actor payoff.

There are other cases where categories are conventional in the sense that they can evolve in multiple ways, even though some of these will be more successful than others. This is still conventionality because there is an arbitrariness to the evolved terms, or a possibility that they could have been otherwise. For example, O'Connor (2013) looks at the emergence of categories in sim-max games where the state space is a line. Actors in these models use

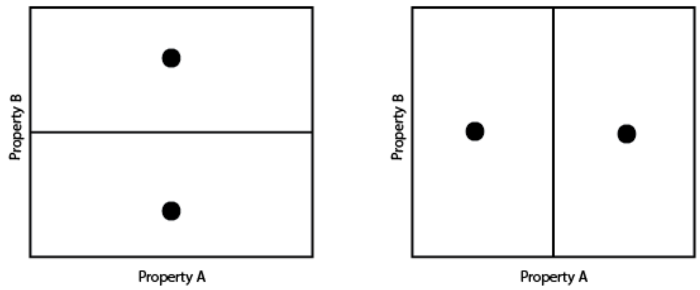



Figure 8: Examples of equally successful Voronoi languages that conventionally divide the state space along different dimensions.

simple learning rules to form categories for the purposes of communication. She finds that while the evolved categories approximate Voronoi languages, there is conventionality as to where category boundaries form. Franke and Correia (to appear) finds similar results for populations that evolve terms under the replicator dynamics (a change rule that is commonly used to represent either biological or cultural evolution). Narens *et al.* (2012) look at the evolution of color terms in a similar model with a circular state space (mimicking human color perception), and find that many different conventional divisions of this space will emerge in learning populations. Figure 9 shows outcomes from a little simulation intended to reinforce this point. I looked at a sim-max game with ten states arrayed on a line. Actors use two signals, and form categories using simple learning rules.[17] Bars represent the proportion

[17]Payoff loss in this game was a gaussian function of Euclidean distance with height two and width four. Actors updated strategies using the basic Roth-Erev reinforcement learning rule (Roth and Erev 1995). (For a detailed, non-technical description of this rule see Barrett (2007).) Each trial involved 1 million rounds of updating. I looked at only 100 trials because the variety of outcomes meant that each one had to be coded by hand. For 35 of these trials, the evolved categories either had vague boundaries, or were non-convex, so I did not include them in this figure. The small number of trials, and missing data, means that these results should not

of simulations where the category boundary ended up at that state (numbers above give the exact number of simulations that evolved to each division). As is clear from the figure, different simulations end up with different categorizations: terms apply to states 1-3 and 4-10, or else states 1-5 and 6-10, etc.

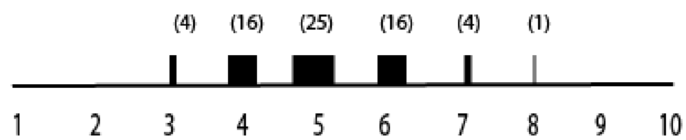



Figure 9: Outcomes from simulations of actors learning to behave in a sim-max game with ten states arrayed on a line and two signals. Bars represent the proportions of simulations that evolved boundaries between terms at each state. Numbers give the exact number of simulations for each outcome.

There is another aspect of conventionality, which has to do with the number of terms employed to cover a space. As mentioned above, cognitive costs usually mean that the number of terms in human languages are limited. But in many cases, a perfectly good language could have more or fewer terms. Again, there will be payoff differences between the languages based on the number of terms, but these might be small. Or the particular demands of a linguistic community may mean that a larger or smaller number of terms is optimal for dividing the space. Figure 10 shows Voronoi languages for the same state space in figure 8, but divided into three and four categories. Again, then, there are multiple possible outcomes of the evolutionary model.

be taken as measuring basins of attraction for this model. They are intended only to show that different outcomes occur across trials, meaning there is conventionality in category boundary.

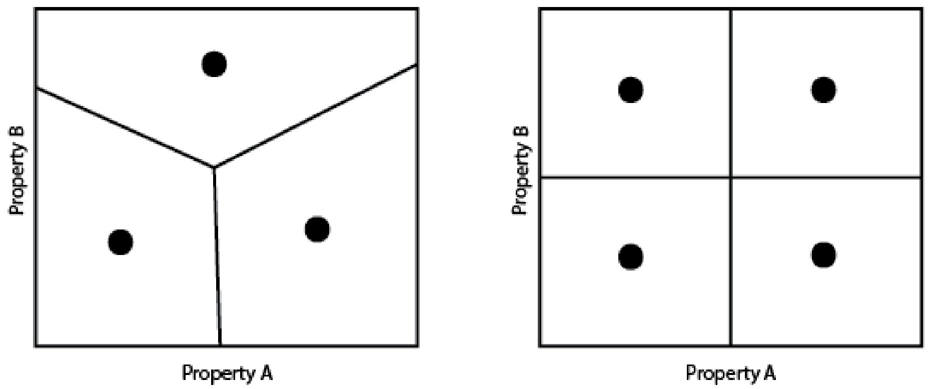



Figure 10: Two Voronoi languages with different numbers of terms.

How do these sorts of conventionality apply to cases with clusters? For each of these sorts of conventionality, we can see the same thing arising in sim-max games where states cluster. Clusters may split along different dimensions. The boundaries between them may be drawn at different places. The number of terms used to cover clusters may be more or less.

In the sim-max game cases just described, the skeptical claims from Barrett (2007) can be reapplied. Actors using these terms might be easily convinced that because they lead to inductive success they must be appropriately tracking real-world features in non-conventional ways.

Returning to possible desiderata for natural kinds terms, the categories just described do not necessarily respect cluster structure. And they certainly violate the possible desiderata that kind terms be fully determined by the world, or completely non-conventional. They are, however, always convex and connected, preserving the, arguably, most important feature of natural terms. In the next section, I will describe the evolution of terms that do not respect convexity and connectedness.

### 4.2 *Functionality and Categories*

The role of inductive success is, of course, key in determining how terms track real world objects. Cluster kinds theorists emphasize this functionality of kind terms. And in section 3 we saw that our evolutionary game theoretic methods confirm that terms ought to evolve to categorize items that, when treated as a kind, generate inductive success.

At this point, though, it will be useful to pull out in more detail just what this means. I have claimed that Voronoi languages will generally, approximately, group similar items together and respect clusterings of items (though they may sometimes fail to do so). I also pointed out, though, that similarity in sim-max games is hashed out in terms of payoff. In other words, when we consider a cluster of items in the state space of a sim-max game, these are items that are similar simply in that they may successfully be responded to in the same way by the relevant actors. This, in itself, tells us nothing about their similarity in property space. In this way, the framework makes salient a dubious assumption underlying some versions of the cluster kinds picture—that inductive success and properties are related in a straightforward way.

One useful way to think about this is as follows. We cannot simply assume that clusters in property space will also be clustered in payoff space. This means that in thinking about the connection between terms and property similarity, we must add an extra translation step to the model, from property to payoff space. Figure 11 shows a representation of how this might look. In this case, we start with a property space where there are three clusters. The translation to payoff space collapses these to the same area of payoff space.

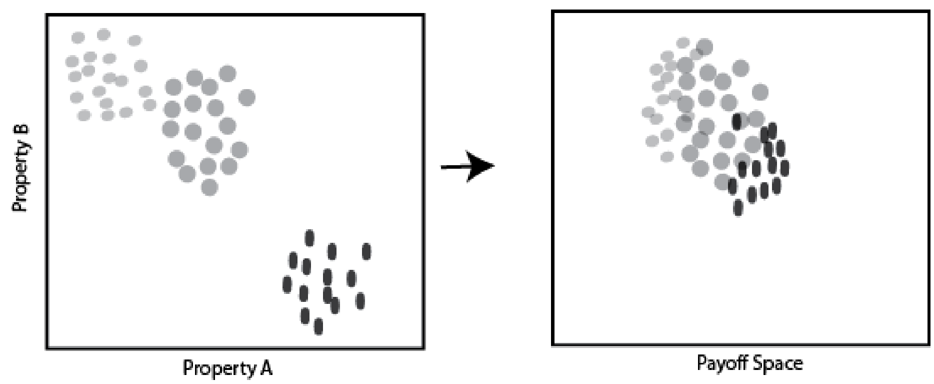



Figure 11: An example of how objects in property space might translate to payoff space.

Of course, given the structure of the world and the needs of a linguistic community, it may be that this translation faithfully preserves the similarity structure of property space. Or, it may be that this translation leads to a payoff structure that is radically different from the underlying property space. To motivate this claim, I will use an example from Millikan (1999)—the chair. Millikan considers this a rough historical kind, which because of recurring facts about human constitution, and copying, will have clusters of similar properties. Alternatively, one might point out that many chairs do not cluster in property space at all, but that they are highly clustered in payoff space for humans who like to sit down. We might say that the pappasan chair, bean bag chair, lazy-boy, the three legged Frank Lloyd Wright chair all are quite different in terms of properties, but very similar in terms of functionality and thus payoffs. We can see figure 11 as a representation of this sort of situation. Suppose that the clusters in the top left of the property space correspond to bean bag chairs and pappasans, and the one in the bottom right to lazy-boys. If these three types of chairs play similar functional roles, they may be translated to a similar area of payoff space for a linguistic community focused on sitting. In what ways does this translation change our picture of how linguistic terms track property clusters? A first stab at answering this question is that anything can happen. If objects from disparate areas of our property space are clustered together in payoff space, they should be grouped under the same category. Or if objects from the same area of property space translate into disparate areas of payoff space, they should be categorized separately.

Of course, this is an overly pessimistic view. As both Boyd and Millikan both emphasize, the fact of property clustering will often be relevant to inductive success. In such cases, we should expect term language to reflect this. What sim-max games give us is a method for predicting when and how terms will reflect property similarity, and the answer is that we should expect them to do so in the cases where property similarity maps to payoff similarity in the right way.

To fully develop the point just made above, there are a few ways that the mapping from property to payoff space might lead to terms that do not respect what we would think of as nature's joints. The first involves categories that move objects close in similarity space apart from each other as a result of the payoff needs of the linguistic community. One example that is somewhat helpful here comes from the literature on color categorization. It is well established that while color categories seem to be influenced by the facts of perceptual color space, they are also highly conventional and differ across societies (Narens *et al.* 2012; Jameson 2010).[18] Berinmo speakers from Papua New Guinea do not distinguish the color categories blue and green typically used by English speakers. Instead, they have terms to track yellowish-green (nol) and blue-ish green (wor) (Roberson 2005). Suggestively, tulip leaves, a favorite vegetable of Berinmo speakers, are green when picked and quickly yellow over time. In other words, the functional needs of the tribe mean that they carefully distinguish between tulip leaves that we would think of as close together in similarity space. Figure 12 shows a representation of what the translation from similarity to payoff space looks like in this case. Tulip leaves that cover a stretched out area of similarity space break into relevantly separate clumps in payoff space. (Presumably these clumps still have a vague boundary, because organic matter decays slowly and by degree.) The terms nol and wor then apply to these payoff clusters.[19]

Another way such distortions happen is when items that we think of as far apart in similarity space show up in the same area of payoff space. The chair example is just such a case. Because the sitting needs of the community are well met by several clusters of items in different areas of property space, we see a linguistic term that groups items that are, in many ways, not property similar. There are plenty such terms in day to day language—dessert, trash, and weapon, for example, are terms that track arguably highly disunified sets of objects

[18] In particular, there is huge variation in the number of basic color categories that different languages employ, as well as variation in the areas of color space that these terms track.

[19] Although this is a case where the linguistic terms are adjectival properties, not kinds terms, we can see how the same thing might happen for kind language as well.

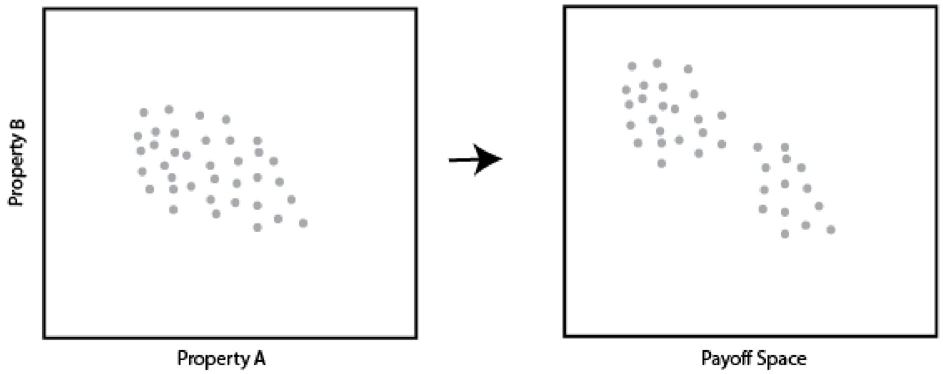



Figure 12: An example of how tulip leaves in property space might translate to payoff space.

that nonetheless play similar functional roles for humans. In each case, treating these objects as a kind leads to inductive success. The result is terms that cover non-connected areas of property space.

## 5 Unnatural Perceptual Categories

One response to the argument just given is that linguistic terms are not solely based on day to day function. In fact, they also rely heavily on the structure of our perceptual categories. Gärdenfors (2000), for example, compellingly argues that most concepts map to convex areas of what he calls conceptual space.[20] A conceptual space is a construct that tracks perceived similarity for an observer. One can generate such a space by having observers compare the similarity of various items and make decisions about which ones are more similar than others,

[20] Relatedly, when it comes to color categories, no language has ever been found that groups two areas of perceptual color space and excludes an area between them (Roberson 2005). In other words, color terms are connected and generally convex in perceptual space.

using whatever criteria they find salient. Once these comparisons are made, one can generate constructs, such as the famous Munsell color space, where distance tracks the stated similarity judgments of the viewer.[21] A convex area of this space will look like the terms of a Voronoi language. The implication is that if terms tend to track concepts, terms also will track perceptual similarity, and not just function, potentially leading to terms that are more likely to be convex in property space.[22]

However, appealing to the role of perception in the emergence of linguistic terms will not completely still worries about unnatural kind terms. Perceptual categories themselves, arguably, do not track real world property structures in a straightforward way. Perception, like language, involves categories that allow actors to facilitate their, either learned or evolved, behavior in the world. Perceptual categories are useful for action because without them it would be impossible for organisms to generalize their behavior appropriately. For example, if an organism, upon encountering a poisonous caterpillar, only learned not to eat that exact sort of caterpillar—exact same weight, coloring, furriness, length—it would learn extremely slowly. In fact, the organism would likely never have the opportunity to reapply the learned lesson. Instead, learning organisms apply what they learn over categories of objects they encounter.[23] Given the similar function of perceptual and linguistic categories—to group

[21]For more on such spaces and how they can be generated, see Krantz *et al.* (1971).

[22]I can now return to the claim from Jäger (2007) that results from sim-max games support the Gärdenfors (2000) picture of convex conceptual categories. This should only be true if similar perceptual/conceptual states are also payoff similar for the relevant linguistic community.

[23]For more on the role of generalization in animal learning see Ghirlanda and Enquist (2003). For an evolutionary analysis of learning generalization and a discussion of its connection to

objects for the purposes of action—we might expect the problems described above for the naturalness of linguistic categories to apply to perceptual categories as well.

O'Connor (2014) uses the sim-max game framework to make this salient. She points out that rather than interpreting this as a game where two actors try to communicate, one can also interpret the sim-max game as involving a single actor who responds to real world stimuli, where this response is mediated by perceptual experience. Under this interpretation, the state of the world is characterized by the perceptual stimuli received by an organism, or alternatively by the external state leading to that stimuli, the act is interpreted as whatever action the organism takes in response to the state. The signal corresponds to some sort of internal process involving the perceptual experience of the organism that begins with perceptual stimuli and ends with action. This way of interpreting the sim-max game provides some confirmation, through mathematical modeling, that, indeed, perceptual categories should evolve to track payoff similarity of states, rather than directly tracking real world properties. This means that perceptual categories, like linguistic ones, potentially track non-convex, and conventional, areas of property space. Mark *et al.* (2010) engage in a similar project, also employing evolutionary modeling. They look at the evolution of perceptual categories where 'objective' similarity and payoff diverge to show that payoff is the property categories will evolve to track. Human color vision, for example, is notorious in philosophy for posing the problem of metamers—that multiple surfaces with sometimes very different reflectance

perceptual categories see O'Connor (2015b).

profiles will appear identical to a human observer. In such a case, because there was no evolutionary pressure to differentiate these sets of surfaces, we did not evolve to do so.[24] To drive home the point of this section, consider a case from vision researcher Donald Hoffman's work. The jewel beetle has evolved to select mates with large, shiny, brown backs. Upon the introduction of discarded beer bottles into their territory, jewel beetles suffered from a serious threat to their fitness. Males spent all their time attempting to mate with beer bottles, instead of females. In this case, we might say that jewel beetles developed perceptual categories to respond to their environment as a result of their fitness-relevant needs. Larger and browner translated into better for mating. The introduction of beer bottles, however, made abundantly clear to observers that their perceptual categories were tracking fitness needs, rather than any sort of natural kinds. (We can hopefully agree that female jewel beetles and beer bottles do not constitute a natural kind in any traditional sense of the term.)

[24]Perception researchers, though not all of them, have made similar arguments without formal tools. Dale Purves and others who support the 'wholly empirical theory of perception' argue that it is a long evolutionary history of success and failure that determines how organisms come to interpret the various sensory stimuli they receive. For example Purves *et al.* (2011), argue that, '...the basis for what we see is not the physical qualities of objects or actual conditions in the world but operationally determined perceptions that promote behaviors that worked in the past and are thus likely to work in response to current retinal stimuli' (1). Donald Hoffman, a co-author on the Mark *et al.* (2010) paper just discussed, has argued specifically for the non-veridicality of perceptual categories. As he says, 'Fitness not accuracy, is the *objective function* maximized by evolution' (Hoffman 2009, 5). His interface theory of perception holds that as a result the relationship between perceptual categories and the world is analogous to the relationship between a computer icon and the structures it represents.

## 6 The Payoff Relevance of Property Space

At this point, it will be useful to step back and think about how the analyses presented impact cluster kinds views. I have argued that when we model the emergence of terms to represent objects in property space, we see failures to meet the desiderata that we might want for natural kind language—non-conventionality, respect for clustering, and convexity and connectedness. One response might go as follows: we need not be concerned about preserving the naturalness of categories in a full property space that captures every imaginable property. Instead, what we are actually trying to do is preserve the naturalness of categories with respect to some delimited set of properties that we care about.

I motivated the property space introduced in section 2 by appeal to a small set of properties that we, as humans, are used to using when describing objects—size, shape, hardness, surface texture. There are infinitely many properties that we might identify, though, as potential candidates to occupy a dimension of this space (melting point, proportion of gold atoms, flexibility, softness, variation in size over time, saltiness, radioactivity, inertia, and so on). There are further properties we might include like average color appearance to humans, color appearance to red-green color blind humans, ability to stimulate taste buds for sweetness in parrots, climbaility by geckos. One might say that these latter properties are not natural ones, but presumably for each of these there are some facts about the world that they track. There are some facts about molecular configuration that restrict which molecules will bind with sweetness receptors on parrots tongues. There are some sets of surfaces that gecko toes can attach to and others they cannot. So why couldn't they be included as dimensions in a property space? Of course, these properties are picked out by their relevance to the payoffs of organisms. But, arguably, this is also true of all the properties we usually think of as natural ones to act as dimensions in a property space. Why do we care about shape of objects? Because shape tends to be relevant to the ways that we interact with objects. Why do we care about surface reflectance? Because it tends to give us information about the make-up of objects that is relevant to our lives. Why don't we usually worry about the ways that surfaces polarize the light they reflect? Because this is generally irrelevant for our day to day function as organisms who do not perceive polarization.

The point here is that our intuitive notion of property space introduced above already incorporates payoff relevance. There is no reason to expect a canonical way to classify the properties of objects independent from the concerns of those interacting with objects. What this means is that throughout the paper, I have been using a sort of shorthand when I said that there are translations from property space to payoff space that may or may not preserve groupings. This is true, but the property space is not a completely user-independent one and payoff relevance is already creeping into its dimensions—we choose which features of objects we care about to include as dimensions of property space.

This sounds like yet another way in which payoff concerns separate cluster kinds terms from real naturalness. But, in fact, it recovers part of the cluster kinds picture, or else deflates the possibility of identifying naturalness with a capital 'N', and leaves something less demanding in its place. If there is no way to canonically pick out properties in a space, why not pick the ones that matter for payoff and just say that cluster kinds terms respect those ones? (In fact, this picture is close to what Boyd has in mind in at least some of his work on cluster kinds (Boyd 1999, 2000), though is quite far from the picture presented by Millikan (to appear). Magnus (2012) has also defended this sort of 'domain-relative' cluster kind as real.) Return to the chair example. While I have argued that payoff concerns lead to a grouping of unnatural items in this case, I might say instead that there are a set of properties that involve the existence of a horizontal, or partially horizontal, surface at the right height for a human bottom to rest on, enough stability of the material below to support this bottom, and a place for a human back to rest against. 'Chairs' pick out a set of objects that exist in an approximately convex area of the space with these properties. Likewise, 'desserts' pick out a set of objects that exist in an approximately convex area of a property space that has dimensions for edibility and ability to stimulate human sweetness receptors.

The picture of naturalness that arises strongly echos the 'promiscuous realism' introduced by Dupré (1995). In order to determine whether kinds terms are natural, we need to first pick out a set of properties for them to be natural with respect to. It will be possible to choose ones that work. Or we might choose others for other reasons and observe that clusters will not be natural with respect to these. The answer to the question 'is X a natural kind?' must always first specify which things we care about naturalness with respect to. Importantly, this means that the same group of objects could be a natural property cluster kind with respect to some set of properties, and unnatural with respect to others.

We might use this picture to say something about natural kinds terms in science. There have been many terms that hooked onto inductively successful sets of objects, for the relevant science, which later were rejected as unnatural. Consider the term 'reptile', which traditionally included crocodiles, lizards, and amphibia, but not birds. This term was later recognized as one that fails to respect phylogenetic groupings. We might say that on certain properties, those relevant to early taxonomists, reptiles picked out an approximately convex area of property space. They do not correspond to the right area of phylogenetic space, and so if this is the relevant property for the sciences, they are rejected as a kind. What changed here was the salient properties of the objects of science for scientists, not the objects themselves. Or consider 'cancer'. This condition has successfully been used as a kind, for the purposes of medicine, because it can often be treated similarly. Likewise, there are sets of properties for which 'cancer' will pick out an approximately convex category. There are also sets of properties for which it will not, and sometimes this means that treatments will not be generalizable across cancers, pushing us towards new terminology that breaks this 'kind' apart.[25]

At this point, I want to make clear that my purposes here are neither to argue for a natural view of cluster kinds, nor to argue against one. Cluster kinds theorists, as I pointed out, will usually be able to find properties for which terms will be natural, if they wish to. And though

[25] See Khalidi (2013) and Plutynski (forthcoming) for philosophical work on cancer as a kind.

these properties may sometimes strike us as themselves ‘unnatural’, payoff concerns are always at play when we choose properties to focus on. And so they may use something like promiscuous realism to defend the naturalness of cluster kinds. On the other hand, for any kind term we can find properties that it will not be natural with respect to, often even properties that we think of as traditionally natural properties. And for any set of properties, we will usually be able to find kinds terms that do not respect them.

The important take-away from the analysis of sim-max games presented here is not direct support for nor evidence against the naturalness of cluster kind terms. Instead, it is a deeper understanding of the conditions under which terms will be natural with respect to which properties, and a better picture of the assumptions underlying claims that terms will track cluster kinds.

## 7 Conclusion

Cluster kinds theorists posit that there exist sets of objects that cluster in property space as a result of historical or homeostatic processes. A result of this clustering is that treating these objects as a kind will lead to inductive success in everyday life, or in the sciences. This paper used evolutionary game theoretic models of the evolution of categories and terms to assess whether and when terms will, in fact, track such clusters. As I show, there are many cases where they will not—as a result of both the conventionality and the functionality of linguistic terms. The role of perceptual categories in term formation does not necessarily help, as perceptual categories are themselves conventional and functional.

Taking the framework seriously, however, presents the possibility of a promiscuously real view of natural kinds. The properties that we choose as natural ones to act as dimensions in a property space are chosen for their relevance to the purposes and activities of humans. Given this, one might say that cluster kinds terms are natural on some property dimensions, and not natural on others. In particular, the framework here gives us a way of predicting whether kinds terms will evolve to respect convex sets in a property space. This will occur when that property space corresponds to the payoff space of the linguistic community in question.

Acknowledgements: Many thanks to Jeffrey Barrett, Brian Skyrms, Kyle Stanford, James Weatherall and audiences at the 8th Quad Fellows conference, LSE Popper seminar, and University of Salzburg for helpful comments on this work. Special thanks to Ruth Millikan for early access to her book manuscript.

Cailin O’Connor

Department of Logic and Philosophy of Science

University of California, Irvine, CA 92697

cailino@uci.edu

## References


Barrett, J. A. [2007]: 'Dynamic Partitioning and the Conventionality of Kinds*', *Philosophy of Science*, **74**(4), pp. 527–546.

Boyd, R. [1989]: 'What realism implies and what it does not', *Dialectica*, **43**(1-2), pp. 5–29.

Boyd, R. [1990]: 'Realism, approximate truth, and philosophical method', *Scientific theories*, **14**, pp. 355–391.

Boyd, R. [1991]: 'Realism, anti-foundationalism and the enthusiasm for natural kinds', *Philosophical studies*, **61**(1), pp. 127–148.

Boyd, R. [1999]: 'Homeostasis, species, and higher taxa', in R. A. Wilson (*ed.*), *Species: New Interdisciplinary Essays*, The MIT Press, chap. 6, pp. 141–185.

Boyd, R. [2000]: 'Kinds as the "workmanship of men": Realism, constructivism, and natural kinds', *Rationalität, Realismus, Revision: Vorträge des*, **3**, pp. 52–89.

Boyd, R. [2010]: 'Homeostasis, higher taxa, and monophyly', *Philosophy of Science*, **77**(5), pp. 686–701.

Dupré, J. [1995]: *The disorder of things: Metaphysical foundations of the disunity of science*, Harvard University Press.

Franke, M. and Correia, J. [to appear]: 'Vagueness and Imprecise Imitation in Signaling Games', *British Journal for the Philosophy of Science*.

Franke, M., Jäger, G. and van Rooij, R. [2011]: 'Vagueness, Signaling and Bounded Rationality', pp. 45–59. Springer.

Gärdenfors, P. [2000]: *Conceptual Spaces: on the geometry of space*, Cambridge, MA: MIT Press.

Ghirlanda, S. and Enquist, M. [2003]: 'A century of generalization', *Animal Behaviour*, **66**(1), pp. 15–36.

Goodman, N. [1972]: 'Seven strictures on similarity', in N. Goodman (*ed.*), *Problems and Projects*, New York: Bobbs-Merrill, pp. 437–446.

Hoffman, D. [2009]: 'The interface theory of perception: Natural selection drives true perception to swift extinction', in S. Dickinson, M. Tarr, A. Leonardis and B. Schiele (*eds*), *Object categorization: Computer and human vision perspectives*, Cambridge, UK: Cambridge University Press, pp. 148–165.

Jäger, G. [2007]: 'The evolution of convex categories', *Linguistics and Philosophy*, **30**, pp. 551–564.

Jäger, G., Metzger, L. P. and Riedel, F. [2011]: 'Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals', *Games and Economic Behavior*, **73**(2), pp. 517–537.

Jameson, K. A. [2010]: 'Where in the World Color Survey is the support for the Hering primaries as the basis for color categorization', *Color Ontology and Color Science*, pp. 179–202.

Khalidi, M. A. [2013]: *Natural categories and human kinds: Classification in the natural and social sciences*, Cambridge University Press.

Krantz, D., Luce, D., Suppes, P. and Tversky, A. [1971]: 'Foundations of measurement, Vol. I: Additive and polynomial representations', .

Lewis, D. K. [1969]: *Convention*, Cambridge, MA: Harvard University Press.

Lipman, B. L. [2009]: 'Why is language vague?', *Boston University*.

Magnus, P. [2012]: *From Planets to Mallards: Scientific Enquiry and Natural Kinds*, Palgrave Macmillan.

Mark, J. T., Marion, B. B. and Hoffman, D. [2010]: 'Natural selection and veridical perceptions', *Journal of Theoretical Biology*, **266**(504-515).

Millikan, R. [to appear]: *Beyond Concepts: Unicepts, Language, and Natural Information*, Oxford University Press.

Millikan, R. G. [1999]: 'Historical kinds and the "special sciences"', *Philosophical Studies*, **95**(1), pp. 45–65.

Narens, L., Jameson, K. A., Komarova, N. L. and Tauber, S. [2012]: 'Language, categorization, and convention', *Advances in Complex Systems*, **15**(03n04), pp. 1150022.

O'Connor, C. [2013]: 'The Evolution of Vagueness', *Erkenntnis*, pp. 10.1007/s10670–013–9463–2.

O'Connor, C. [2014]: 'Evolving Perceptual Categories', *Philosophy of Science*.

O'Connor, C. [2015a]: 'Ambiguity Is Kinda Good Sometimes', *Philosophy of Science*, **82**(1), pp. 110–121.

O'Connor, C. [2015b]: 'Evolving to Generalize: Trading Precision for Speed', *The British Journal for the Philosophy of Science*.

Plutynski, A. [forthcoming]: *Explaining Cancer: Philosophical Issues in Cancer Causation, Evidence and Explanation*, Oxford University Press.

Purves, D., Wojtach, W. T. and Lotto, R. B. [2011]: 'Understanding vision in wholly empirical terms', *Proceedings of the National Academy of Sciences*, **108**(Supplement 3), pp. 15588–15595.

Roberson, D. [2005]: 'Color categories are culturally diverse in cognition as well as in language', *Cross-Cultural Research*, **39**(1), pp. 56–71.

Roth, A. E. and Erev, I. [1995]: 'Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term', *Games and Economic Behavior*, **8**, pp. 164–212.

Weibull, J. W. [1997]: *Evolutionary game theory*, MIT press.