Dynamic Partitioning and the Conventionality of Kinds*

Jeffrey A. Barrett†‡

Lewis sender-receiver games illustrate how a meaningful term language might evolve from initially meaningless random signals (Lewis 1969; Skyrms 2006). Here we consider how a meaningful language with a primitive grammar might evolve in a somewhat more subtle sort of game. The evolution of such a language involves the co-evolution of partitions of the physical world into what may seem, at least from the perspective of someone using the language, to correspond to canonical natural kinds. While the evolved language may allow for the sort of precise representation that is required for successful coordinated action and prediction, the apparent natural kinds reflected in its structure may be purely conventional. This has both positive and negative implications for the limits of naturalized metaphysics.

1. Natural Kinds. Philosophers have often supposed that the metaphysical structure of the world is somehow reflected in the structure of language. Indeed, much of contemporary metaphysics arguably consists in inferring metaphysical morals from the structure of ordinary language and our intuitions concerning its proper use. While naturalists have traditionally been skeptical of drawing metaphysical conclusions from ordinary language, they have nevertheless often been willing to take the descriptive language of our best scientific theories as a guide to reliable metaphysics. This is a consequence of the naturalist's epistemic commitment to the truth of our best theories. As Quine put the naturalistic commitment:

> The philosophical naturalist begins his reasoning within the inherited

*Received February 2007; revised September 2007

†To contact the author, please write to: Department of Logic and Philosophy of Science, School of Social Sciences, 3151 Social Science Plaza A, University of California, Irvine, CA 92697-5100; e-mail: jabarret@uci.edu.

‡I would like to thank Brian Skyrms, Simon Huttegger, Rory Smead, Kevin Zollman, and Samuel Park for helpful comments and discussions. I would also like to thank the referees who looked at this paper for their excellent suggestions.

Philosophy of Science, 74 (October 2007) pp. 527–546. 0031-8248/2007/7404-0005$10.00

> world theory as a going concern. He tentatively believes all of it, but believes also that some unidentified portions are wrong. He tries to improve, clarify, and understand the system from within. (Quine 1981, 72)

For this sort of naturalist, ontological questions are on a par with questions of natural science; and, for Quine, at least, tentatively believing the inherited world theory involves tentatively believing in the existence of the kinds of objects quantified over in our best empirical theories (Quine 1953, 45). The naturalist might thus suppose that what objects and kinds there are is something that one might infer from the structure of our best descriptive language as it is used in our best empirical theories.

Insofar as our best empirical theories are epistemically preferable to our everyday commonsense descriptions of the world, this approach might be expected to be more reliable than trying to infer the metaphysical structure of the world from ordinary language. But the philosophical naturalist presumably does not want to recapitulate the methods of ordinary language metaphysics at the level of description of our best theories. Supposing that one can infer the metaphysical structure of the world from the structure of our best theoretical language might well be a mistake that differs only in degree from supposing that one can infer the metaphysical structure of the world from the structure of ordinary language. If so, then what might the naturalist reliably infer about the world from the language of our best descriptive theories? One way to approach this question is by considering how our best descriptive language might have evolved. Knowledge of its possible source might then shed light on its relation to the world.

Nelson Goodman held that descriptive language evolves to include those natural-kind predicates that in fact develop a track-record of successful application in inductive inferences. More specifically, our descriptive kind language evolves as successful projections lead to better entrenched kind predicates, where a hypothesis is projected when it is adopted after some but not all of its instances have been examined and determined to be true and a better entrenched predicate is one that has in fact been successfully used in past inductive projections. The predicate "is green" is better entrenched when an hypothesis like "All emeralds are green" is adopted and confirmed on some subset of possible confirming evidence; and the advantage of "is green" over its rival "is grue" is that the former has become entrenched through successful inductive projections and hence is better suited for use in future hypotheses (Goodman

1965, 87 and 94).[1] Goodman further believed that the degree to which a kind term is entrenched might allow one to distinguish between "genuine" and "artificial" kinds: "For surely the entrenchment of classes is some measure of their genuineness as kinds; two things are the more akin according as there is a more specific and better entrenched predicate that applies to both" (123).

In broad terms, Quine endorsed Goodman's account of the evolution of descriptive language and the evolution of natural-kind terms more specifically:

> We revise our standards of similarity or of natural kinds on the strength, as Goodman remarks, of second-order inductions. New groupings, hypothetically adopted at the suggestion of a growing theory, prove favorable to inductions and so become "entrenched." We newly establish the projectibility of some predicate, to our satisfaction, by successfully trying to project it. In induction, nothing succeeds like success. (Quine 1969, 165)

But in order to track successful inductive projections, Quine believed that one must have an innate sense of kinds. He further held that "a sense of similarity or of kinds is fundamental to learning in the widest sense--to language learning, to induction, to expectation" (1969, 166). If this is right, then a sense of at least some kinds is not learned but is a precondition for learning at all.

That the use of a particular kind term in a successful inductive inference should be expected to increase the likelihood of its use in other inductive inferences is certainly plausible, but one might also suspect that less is required for the successful evolution of a descriptive kind language than second-order induction grounded in an innate sense of kinds, at least insofar as one understands this characterization of the preconditions for the evolution of language as one that presupposes that one already has a language of kinds in which to reason. Perhaps better, whether the evolution of a successful descriptive language of kinds requires second-order induction grounded in an innate sense of kinds depends on what one means. Here we will consider how a successful descriptive kind language

1. In this sense, Goodman's new problem of induction reduces to an instance of the old problem of induction, which Goodman believed was resolved by evolving an inductive logic that in fact represents those arguments we are willing to accept given what we learn as we evolve the logic. For Goodman, then, our theories, descriptive language, and logical rules are all subject to evolutionary revision with an aim of reflective equilibrium. While Quine perhaps focused more on the accommodation of empirical evidence than on the evolution of reflective equilibrium in judgment, there is clearly significant overlap between Quine's and Goodman's evolutionary views.

might evolve given the modest resources of a sender-receiver signaling game.

Lewis (1969) introduced sender-receiver games as a way of investigating how successful linguistic conventions might evolve from initially random signaling. We will first consider basic Lewis signaling games, which allow for the evolution of term languages; then we will consider syntactic games, extensions of the Lewis signaling game that are sufficiently rich to co-evolve a kind language and a corresponding systematic partition of the state space. While there are preconditions for the evolution of such descriptive languages, they are relatively weak. And while there are conclusions concerning the nature of the world that one might reliably draw from the structure of such an evolved descriptive language, they are also relatively weak. Perhaps too weak to qualify as metaphysical conclusions.

2. Lewis Signaling Games. A basic Lewis signaling game has two players: the sender and the receiver. In an n-state/n-term signaling game there are n possible states of the world, n possible terms the sender might use as signals, and n possible receiver actions, each of which corresponds to a state of the world. Nature chooses a state at random on each play of the game. The sender then observes the state and randomly sends a term to the receiver, who cannot directly observe the state of the world. The receiver observes the term then randomly chooses an act. If the receiver's action matches the state of the world, the play counts as a success; if not, it counts as a failure.

The sender and receiver may learn from their record of successes and failures on repeated plays of the game. Whether and how quickly they learn will depend on their learning strategy, where a learning strategy here is just a disposition for differentially updating their random signal and action dispositions given their track record of successes and failures. If the sender and the receiver evolve to a state where their signals lead to actions that are more successful than chance, then they have evolved a more or less efficient language. The efficiency of the evolved language can be measured by the expected signal success rate (the expected ratio of successful actions to the number of plays given the sender's and receiver's current dispositions) or by the mean information content of a signal (where $\log_2(n)$ bits is sufficient to specify a particular state from among n possible states). Lewis called a system that evolves to a maximally efficient language a *signaling system*. For a perfect signaling system in a Lewis signaling game each state of the world corresponds to a term in the language and each term corresponds to an act that matches the state of the world, so each signal leads to a successful action. For the 2-state/2-term game (see Figure 1), a perfect Lewis signaling system would have

Figure 1.—A basic 2-state/2-term Lewis signaling game.

an expected signal success rate of 1.0 and each signal would communicate one bit of information.

We will begin by considering a basic 2-state/2-term Lewis signaling game with simple reinforcement learning. Here there are two possible states of the world (*A* and *B*), two possible terms (*0* and *1*), and two possible acts (*A* and *B*), each of which is successful if and only if the corresponding state of the world obtains. The sender has an urn labeled *state A* and an urn labeled *state B*, and the receiver has an urn labeled *signal 1* and an urn labeled *signal 2*. The sender's urns each begin with one ball labeled *signal 1*and one ball labeled *signal 2*, and the receiver's urns each begin with one ball labeled *act A* and one ball labeled *act B*.

On each play of the game the state of the world is randomly determined with uniform probabilities; then the sender consults the urn corresponding to the current state and draws a ball at random with uniform probability for each ball. The signal on the drawn ball is sent to the receiver. The receiver then consults the receiver urn corresponding to the signal and draws a ball at random. If the action on the drawn ball matches the current state of the world, then the sender and the receiver each return their drawn ball to the respective urn and add another ball to the urn with the same label as the drawn ball; otherwise, the sender and receiver just return their drawn ball to the respective urn. On the basic urn learning strategy, there is no penalty for the act failing to match the state. The game is repeated with a new state of the world.

Simple reinforcement learning has a long psychological pedigree. The urn learning strategy described above models Richard Herrnstein's (1970) *matching law* account of choice. On Herrnstein's account, the probability of choosing an action is proportional to the accumulated rewards, which in urn learning is represented by the accumulated biases in the sender's and receiver's dispositions to signal and to act. Herrnstein's *matching law* was itself a quantification of Thorndike's *law of effect* (1898) for the conditioning of stimulus response relations by experience. Simple reinforcement learning continues to be used as a first approximation to animal and human learning.[2]

2. Simple reinforcement learning has been used by Roth and Erev (1995) to model experimental human data on learning in games, by Skyrms and Pemantle (2000) to

The basic 2-state/2-term signaling game with simple reinforcement learning presents an apparently difficult context for the evolution of a successful language. The space of possible states is symmetric with no special saliencies, and the strategy for updating dispositions here is among the simplest possible learning strategies. One might imagine that if a successful term language can evolve in this context, then it is all the more plausible that a successful language might similarly evolve in contexts where there are special saliencies or more sophisticated learning strategies. The relative difficultly of alternative evolutionary contexts, however, is often counterintuitive.[3]

Skyrms (2006) has shown for simple reinforcement learning and Huttegger (2007a) has shown for the replicator dynamics, that signaling systems always evolve in the 2-state/2-term signaling game with evenly distributed states of nature.[4] Skyrms (2006) has also shown that signaling systems always evolve for two senders and one receiver when the senders observe different, prearranged partitions of the state space. In the case of simple reinforcement learning it is easy to get a sense of how this works. Adding balls to the signal and act urns when the act is successful changes the relative proportion of balls in each urn, which changes the conditional probabilities of the sender's signals (conditional on the state) and the receiver's acts (conditional on the signal). The change in the proportion of balls of each type in each urn increases the likelihood that the sender and receiver will draw a type of ball that will lead to successful coordinated action. Here the sender and receiver simultaneously evolve and learn a meaningful language. That they have done so is reflected in their track-record of successful action.

---

model social network formation, and by Skyrms (2004, 2006) to model learning in the context of Lewis signaling games. Huttegger (2007a, 2007b) has also studied 2-state/2-term Lewis signaling games as models for the evolution of term languages for populations using the closely related replicator dynamics (see note 4).

3. While an even distribution of states may seem to contribute to a difficult environment for language evolution, it turns out that, for example, it is harder for perfect signaling to evolve under simple reinforcement learning when the probability distribution over states of the world is not uniform. In this case, at least, the state asymmetry hurts rather than helps in evolving a perfect language. What happens here is that the agents may get a high enough success rate by always associating every available term in their language with the more likely states and ignoring the less likely states in their language; and since there is no punishment for failure on this learning strategy, there is no evolutionary pressure to undo these reinforced dispositions. See Huttegger (2007a) for more details.

4. The replicator dynamics updates the ratios of strategy types in a population of agents by increasing the representation of successful types in subsequent generations in a way that is analogous to how simple reinforcement learning updates a single agent's dispositions.

TABLE 1. Run Failure Rates for Lewis Signaling Games with Urn Learning.

Model	Run Failure Rate
3-state/3-term	.096
4-state/4-term	.219
8-state/8-term	.594

The situation is more complicated for Lewis signaling games with more (or fewer) states or terms or if the distribution of states is biased (see Barrett 2006 and Huttegger 2007a). In such modified games, suboptimal equilibria may develop and prevent uniform convergence to perfect signaling. Table 1 shows success rates for Lewis signaling games with more than two states and terms (see Barrett 2006 for more details). Here there are 10^3 runs of each model with 10^6 plays/run. If the signal success rate is less than 0.8, then the run is taken to fail.[5]

While these results illustrate failures in uniform convergence to perfect signaling, each system is always observed to do better than chance on the simulations and hence to evolve a more or less effective language. In those cases where perfect signaling fails to evolve in the 3-state/3-term game, the system approaches a signaling success rate of about 2/3, where one would expect a chance signal success rate of 1/3.[6] Similarly, in the 4-state/4-term game, when a system does not approach perfect signaling, it nevertheless approaches a success rate of about 3/4.

The behavior of the 8-state/8-term system is more complicated since there are several partial pooling equilibria corresponding to different signal success rates (with a different likelihood for each). The distribution of signal success rates in the 8-state/8-term game with 10^3 runs and 10^6 plays/run is given in Table 2.

The partial pooling equilibria that limit convergence to perfect signaling in these systems are the result of simple reinforcement learning. If one

5. The magnitude of the initial dispositions and of the reinforcements makes a difference to how such systems evolve. In some games, for example, one gets much better learning with simple reinforcement if one starts with initial dispositions that are small relative to the magnitude of the reinforcements. Here it is assumed that the magnitude of the reinforcements of the dispositions is equal to the magnitude of the initial disposition associated with each possible action. For simple urn learning this is represented by each urn starting with one ball of each possible action type and being updated with one ball on success.

6. Systems that approach a signaling success rate of 2/3 here do not learn to signal reliably with two out of three terms; rather, such systems approach a partial pooling equilibrium where two of the terms correspond to the same state-act pair and the other term is used to represent both of the other state-act pairs, and the sender and the receiver follow (different) mixed strategies. See Barrett (2006) for more details.

TABLE 2. Distribution of Signal Success Rates in the 8-state/8-term Signaling Game.

Signal Success Rate Interval	Proportion of Runs
[.0, .50)	.000
[.50, .625)	.001
[.625, .75)	.045
[.75, .875)	.548
[.825, 1.0]	.406

allows for both positive reinforcement on success and negative reinforcement on failure, as an example of a slightly more sophisticated learning strategy, then one typically observes better convergence to perfect signaling systems. On the 8-state/8-term $(+2, -1)$ signaling game, success is rewarded by adding to the relevant urns two balls of the type that led to success and failure is punished by removing from the relevant urns one ball of the type that led to failure. As illustrated in Table 3, this learning strategy more than doubles the chance of perfect signaling evolving in the 8-state/8-term game. And more sophisticated learning strategies do better yet.[7]

The upshot is that while perfect signaling does not always evolve in basic Lewis signaling games, it very often does, even in the context of some of the simplest learning strategies one might imagine.

3. Ewok Kinds. Consider the status of such an evolved language. Suppose that there are two types of state that are salient to Ewoks and that they have consequently evolved a primitive but successful term language in a way that is well-modeled by a 2-state/2-term Lewis sender-receiver game. More specifically, suppose that Ewoks evolve to signal "yub-nub" in time of peace and "glo-wah" in time of war, and that each signal nearly always leads to appropriate actions in Ewoks who hear it. Having evolved Ewokian (their primitive but successful language), Ewoks prosper.

It is sufficient for the evolution of such a descriptive language here that: (i) the sender's signals can be individuated by both the sender and receiver; (ii) the sender's signals can be influenced by the state; (iii) the receiver's actions can be influenced by the sender's signals; and (iv) the sender and

7. See Barrett (2006) for examples of other learning strategies. Note, however, that sender-receiver games need not always approach perfect signaling in order to faithfully model the evolution of natural language insofar as natural language rarely allows for perfect signaling. Indeed, one might imagine the evolution of our best descriptive languages as evolution through successive suboptimal equilibria. Such a view might help to explain the punctuated-equilibrium character of scientific progress in description.

TABLE 3. Distribution of Signal Success Rates in the 8-state/8-term (+2, −1) Signaling Game.

Signal Success Rate Interval	Proportion of Runs
[.0, .50)	.000
[.50, .625)	.000
[.625, .75)	.002
[.75, .875)	.110
[.825, 1.0]	.888

receiver update their conditional dispositions, the sender's dispositions conditional on the state and receiver's dispositions conditional on the signal, in a way that privileges a particular map from states to actions. No special second-order induction over successful projections or innate sense of kinds is required beyond the sender and receiver having effective dispositions for updating their signal-act dispositions.[8]

This last point is important. While one might be tempted by their impressive track record of successful coordinated action to imagine that Ewoks have subtle mental lives, the story of the evolution of Ewokian just concerns how signal-act dispositions are updated. A state stimulates a random signal from the sender, then the signal stimulates a random action in the receiver. If the action matches the state, then this stimulates the signal-act dispositions to be updated to make the sender's last signal more likely given the last state and the receiver's last action more likely given the last signal. And all that it means for an action to match the state is that it is an action that in fact leads to positive reinforcement of the signal-action dispositions that led to it. In this sense, the pairing of states and actions that defines success in their signaling game simply reflects a feature of the way Ewoks are built. Explaining how they might have evolved to be built this way rather than another would require a different evolutionary story. And explaining how Ewoks may someday evolve to have subtle mental lives that may in turn influence the future evolution of their language would require yet another evolutionary story; in this case, one involving a careful characterization of what it takes to have a subtle mental life.[9]

Since each of term of Ewokian might have evolved to refer to the other state, there is clearly an element of convention in its evolution. But since

8. If we imagine Quine speaking from the perspective of naturalized epistemology as a branch of behavioral psychology, this might be all he meant when he claimed that a sense of similarity or of kinds is fundamental to learning a language.

9. Insofar as we believe that humans have evolved from organisms that did not have subtle mental lives to organisms that do, we believe that such evolutionary stories are possible.

each term of Ewokian corresponds to a kind of state salient to successful Ewok action, one might also argue that there is at least a sense in which the structure of Ewokian tracks natural kinds. The structure of Ewokian, however, is more a reflection of Ewok dispositions than it is a reflection of the structure of the world more generally. Since the kinds that evolve depend on the Ewoks' contingent dispositions for updating their signal-act dispositions, they are presumably not the sort of kinds one would want metaphysics to track.

This is an important point, but it is perhaps also difficult to see here. If Ewoks had started with different dispositions to update their signal-act dispositions, their language might have evolved to carve the world in a very different way. That this contingency of linguistic kinds is not obvious is in part due to the way the game is specified. The specification of the game starts with the observation that there are two types of state that are salient to Ewoks. This fixes the structure of the state space and the map between states and successful actions, and hence constrains the possible languages that might evolve. But what is salient to Ewoks is determined by how they update their dispositions, which is just a function of their contingent nature.

There is also a sense in which the simplicity of the 2-state/2-term game masks the contingent evolution of the language here. Since there are only two types of states in the 2-state/2-term game, there is only one way that the Ewoks might evolve to successfully partition the state space (up to a permutation of terms); hence, this one way looks canonical. The apparent reflection of the structure of the world in the structure of Ewokian then is baked into the specification of the evolutionary game. In order to clearly see how the structure of the evolved language may be contingent, one must consider more subtle evolutionary games.

4. Dynamic Partitioning of States and Syntactic Games. If there are more states and more corresponding acts than available terms in a signaling game, then it is impossible to evolve a term language that allows for perfect signaling, since there are not enough terms to represent each state-act pair. Skyrms (2006) has shown that a language may evolve in the context of such informational bottlenecks to do as well as possible given the available linguistic resources.[10] The language achieves this partial degree of expressive success by co-evolving with a coarse-grained partition of the state space that best fits the available linguistic resources under the

10. In the 3-state/2-term signaling game, for example, the best possible signal success rate given the linguistic resources available is 2/3. With 10^6 plays/run and 200 runs, the signal success rate is always found to approach this best possible rate (Barrett 2006).

demand of successful action. In a 3-state/2-term signaling game, for example, one term might evolve to correspond to states *A* and *B*, for example, and the other to state *C*. The extra degree of freedom in the relationship between states and language here allows one to see a significant way in which evolved languages may be conventional: different languages may be equally successful but partition the state space differently. In the 3-state/2-term game, the first term might evolve to correspond to states *A* and *C* and the second to state *B*. And unlike the different permutations of terms that might evolve in a 2-state/2-term game, term languages that evolve in the context of an informational bottleneck are typically not directly intertranslatable.

One might at first suspect that such conventional dynamic partitioning is an artifact of insufficient linguistic resources for evolving perfect signaling. We will see, however, that even languages that allow for perfect signaling may exhibit this sort of conventionality in their evolution. Indeed, there is good reason to expect dynamic partitioning to be ubiquitous in the evolution of language. In particular, conventional dynamic partitioning might be expected as a feature of the evolution of any language with a nontrivial grammar.[11]

While the signaling games considered so far illustrate how a simple term language might evolve from random signaling, one might also model the evolution of more subtle linguistic conventions in the context of signaling games. Syntactic games are an extension of basic Lewis signaling games where there are more states relevant to successful action than available terms, but also where there is more than one signal available to send on each play of the game. The new syntactic degree of freedom may then evolve to be exploited for the representation of states. Syntactic games involve a more subtle sort of dynamical partitioning than the 3-state/2-term signaling game just considered. In a syntactic game, a language may evolve that allows for perfect signaling using what appear to be systematically interrelated natural kinds.

A 4-state/2-term/2-sender syntactic game has two senders who observe the state of the world, then each sends a signal of either *0* or *1* to a single receiver (see Figure 2). Each signal is independent in the sense that neither sender knows what the other sent. The receiver knows each signal and which sender sent it, which might be thought of as the order in which the terms are sent, but does not know the state of the world. There are four possible receiver acts, each corresponding to one of the four states.

11. In a simple bottleneck game, one might argue that there is a correspondence between kind terms in the language and *coarse-grained* natural kinds in the world. Such an argument is typically not available in the case of a language that evolves in a syntactic game.

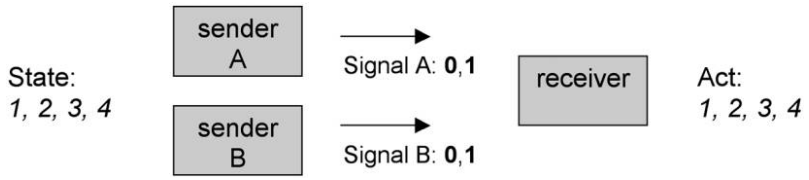

Figure 2.—A 4-state/2-term/2-sender syntactic game.

A receiver's act is successful if and only if the corresponding state obtains. We will start by assuming simple reinforcement learning where each sender has four urns (one for each possible state) and the receiver has four urns (one for each possible combination of two signals). The two senders and the receiver add a ball of the successful signal or act type to the appropriate urn on success and simply replace the drawn balls on failure. We will also suppose uniformly distributed states of nature.

On the face of it, this is a very difficult context for the evolution of language. Since there are four states and four acts but only two terms, a signaling system can evolve here only if the senders and receiver learn to use the available syntactic structure to code for the four state-act pairs. Further, as in the Lewis signaling games considered earlier, the state space is symmetric with no special saliencies and the learning dynamics allows for only positive reinforcement. And since neither sender knows what signal was sent by the other, senders cannot directly learn to correlate their signals to successfully code for the state. Nevertheless, the senders and receiver typically evolve a perfect language that codes for each of the four state-act pairs in this game.

As with the 4-state/4-term Lewis signaling game discussed earlier, the 4-state/2-term/2-sender syntactic game with basic urn learning approaches perfect signaling approximately 3/4 of the time. Table 4 shows results of simulations of the 4-state/2-term/2-sender game with a comparison to the results of the basic 4-state/4-term Lewis signaling game for comparison.

On a successful run of the 4-state/2-term/2-sender game, the senders and receiver simultaneously evolve coordinated partitions of the state space and a code where the senders' signals, taken together, select a state in the partitions. The code that evolves on a successful run is a permutation of "**00**" representing state *1*, "**01**" representing state *2*, "**10**" representing state *3*, and "**11**" representing state *4*. Partial pooling equilibria are responsible for those runs where perfect signaling does not evolve. Such failed runs are observed to approach a signaling success rate of about 3/4, and thus still do better than chance and, in this sense, represent the evolution of an imperfect language.

TABLE 4. SUCCESS RATES FOR THE 4-STATE/2-TERM/2-SENDER SYNTACTIC GAME (the 4-state/4-term Lewis Signaling Game Is Included for Comparison).

Number of Plays/Run	4-State/2-Term/2-Sender Game Run Success Rate (> .8 Signal Success Rate)	4-State/4-Term Game Run Success Rate (> .8 Signal Success Rate)
10^6	.731 [2000 runs]	.781 [1000 runs]
10^7	.75 [100 runs]	.83 [100 runs]
10^8	.73 [300 runs]	.81 [100 runs]

More complex coding schemes evolve in syntactic games with more states and more senders. In the 8-state/2-term/3-sender syntactic game, there are eight states and three independent senders, each restricted to the terms *0* and *1*. With basic urn learning, this model approaches a perfect signaling system about 1/3 of the time. In this case, each of the eight possible states of the world are represented by a three-term binary string. The distribution of signal success rates for 10^3 runs with 10^6 plays/run is given in Table 5 and the corresponding Lewis signaling game (with one sender and eight terms) is included for comparison.

Again, while the evolution of suboptimal equilibria sometimes prevents the evolution of perfect signaling, a more-or-less effective language always evolves.

A slightly more sophisticated learning strategy can significantly improve the chances of evolving perfect signaling in a syntactic game. Table 6 gives the results of 8-state/2-term/3-sender $(+3, -1)$ system on 10^3 runs with 10^6 plays/run (the learning strategy is reinforcement where the senders and receiver add three balls of the successful type on success and remove one ball of the failed type on failure). Results for the 8-state/8-term $(+3, -1)$ game (one sender with eight terms and with the same positive and negative reinforcement as the syntactic game) is included for comparison. With its slightly upgraded learning dynamics, the 8-state/2-term/3-sender $(+3, -1)$ system approaches perfect signaling on most runs.

On a successful run of the 8-state/2-term/3-sender game, for example, each sender evolves with the receiver a different but precisely coordinated two-cell partition of the state space where each signal selects a set of four states and the three signals together select a particular state. Table 7 illustrates how the partitions might evolve on a particular run. Here if A sends **1,** B sends **0,** and C sends **1**, then state *4* obtains

The co-evolution of a partition of the state space and the meaningful exploitation of available linguistic structure here provides an example of how a relatively subtle kind language might evolve in the context of modest evolutionary resources. Each sender's signal selects a kind of state, and the specification of three kinds of state together selects a single state at the level of individuation required for successful action given the dis-

TABLE 5. Distribution of Signal Success Rates for the 8-State/2-Term/3-Sender Syntactic Game (the Distribution for 8-State/8-Term Games Is Included for Comparison).

Signal Success Rate Interval	8-State/2-Term/3-Sender Game Proportion of Runs	8-State/8-Term Game Proportion of Runs
[.0, .50)	.000	.000
[.50, .625)	.001	.001
[.625, .75)	.081	.045
[.75, .875)	.589	.548
[.825, 1.0]	.329	.406

positional payoffs of the game. While uniformly successful action might lead one to suspect that the structure of the evolved language reflects canonical natural kinds, there is nothing special about the evolved partitions of the state space here beyond the fact that they work well for the purposes at hand, which, again, just means that the agents' dispositions to send and act have evolved to a reflective equilibrium with the way that they update these dispositions.[12]

5. Wookiee Kinds. Suppose that there are eight types of state salient to successful Wookiee action, and that Wookiees have evolved a language in a way that is well-modeled by the 8-state/2-term/3-sender syntactic game. While each effective partition of the state space is equally likely, suppose that Wookiees in fact evolve to say "blue" when the state is *1*, *5*, *8*, or *2* and "green" when it is *3*, *7*, *4*, or *6*; "hot" when the state is *1*, *3*, *8*, or *4* and "cold" when it is *5*, *7*, *2*, or *6*; and "odd" when the state is *1*, *5*, *3*, or *7* and "even" when the state is *2*, *8*, *4*, or *6*. In Wookieespeak, then, "blue cold even" means that state *2* obtains; which may, for example, lead the listener to take a large umbrella to the picnic. That Wookieespeak has a different linguistic structure than Ewokian is not a consequence of Wookiees inhabiting a different world with a different canonical metaphysical structure; rather, it is a consequence of Wookiees having different dispositions for updating their signal-act dispositions than Ewoks—dispositions better represented by an 8-state/2-term/3-sender syntactic game than by a 2-term/2-state sender-receiver game. And in the happy state of having evolved a perfectly effective language for their purposes, the Wookiees enjoy successful Wookiee action at every turn.

Some of the more speculative Wookiees, however, are not satisfied simply to have evolved a language that allows for successful coordinated action. Perhaps tempted by the apparent systematic fit between their language and the world suggested by the track-record of its successful use,

12. See Barrett (2007) for additional details concerning the evolution of syntactic games.

TABLE 6. Distribution of Signal Success Rates for the 8-State/2-Term/3-Sender (+3, −1) Syntactic Game (the Distribution for the 8-State/8-Term [+3, −1] Is Included for Comparison).

Signal Success Rate Interval	8-State/2-Term/3-Sender (+3, −1) Proportion of Runs	8-State/8-Term (+3, −1) Proportion of Runs
[.0, .50)	.000	.000
[.50, .625)	.000	.000
[.625, .75)	.004	.004
[.75, .875)	.157	.225
[.825, 1.0]	.839	.771

they imagine that it is possible to infer the metaphysical structure of the world from the structure of their language. More specifically, they conclude that there are three canonical qualities: *color* (*blue* or *green*), *temperature* (*hot* or *cold*), and *parity* (*odd* or *even*). While they are pleased with themselves at successfully carving nature at its joints, there do remain points of contention inviting further speculation. Some claim that it is clear and distinct that *color*, *temperature*, and *parity* suffice to fully characterize the essential attributes of all matter while others claim that it is certain there are other possible qualities of matter, though perhaps only realized in counterfactual worlds. Some believe that the discovered qualities are not attributes of matter at all but rather attributes of appearances. Others concede this point but argue that the qualities correspond to forms of sensible intuition or pure concepts that serve as necessary preconditions for the very possibility of Wookiee experience.

Other, more practical, Wookiees are pleased with their track-record of successful coordinated action but, if possible, would like to do better. While it is unclear to this group why an analysis of the structure Wookieespeak should be expected to teach one something about the canonical structure of the world, they do believe that there are meaningful questions that can be posed in their descriptive language that might lead to a better descriptive language if they can be answered through careful empirical investigation. If this investigation leads them to change how they update their signal-act dispositions, then Wookieespeak may no longer represent a reflective equilibrium. Suppose, for example, that Wookiees turn to consider questions concerning possible *correlations* between the three qualities represented in their language, make whatever empirical observations they take to be relevant, then puzzle over how the observed regularities might be best represented. Since the payoffs of such puzzling may be very different from those of the game in which Wookieespeak originally evolved, the Wookiees may conclude that their original evolved language represents a suboptimal equilibrium (perhaps akin to pooling or bottleneck equilibria).

In any case, if the Wookiees' dispositions to update their signal-act

TABLE 7. EXAMPLE OF THE COORDINATED PARTITIONS THAT MAY EVOLVE ON THE 8-STATE/2-TERM/3-SENDER SYNTACTIC GAME.

Sender A	Sender B	Sender C
Sends 0 on state 1, 5, 8, or 2	Sends 0 on state 1, 3, 8, or 4	Sends 0 on state 1, 5, 3, or 7
Sends 1 on state 3, 7, 4, or 6	Sends 1 on state 5, 7, 2, or 6	Sends 1 on state 2, 8, 4, or 6

dispositions change, then their language will evolve under new constraints. Some of the old language may be preserved, but new terms and syntactic conventions may find use and some of the old terms and systematic conventions may fall out of use in the evolutionary context of the new game. The new evolved language may not partition the world in quite the same way as the original language did. In Newookieespeak *spin*(which comes in varieties *up* and *down*) may, for example, be introduced to represent possible states of the world and *color* terms may no longer partition states as they did in the original Wookieespeak.[13]

Some of the more speculative Wookiees note the changes in the evolved descriptive language, and, attracted by its improved descriptive success, conclude that it is Newookieespeak, not Wookieespeak, that in fact carves nature at its joints. While they are pleased that they can now reliably infer genuine natural kinds from their descriptive language and pity their philosophical colleagues who are still trying to infer metaphysics from an outdated language, there do remain points of contention inviting further speculation. Some believe that the old terms of Wookieespeak that no longer show up in the most basic Newookieespeak state descriptions correspond to natural but emergent qualities that supervene on the more basic natural qualities represented in the new descriptive language, others argue that the old distinctions are part of a folk language that should be discarded with other philosophical nonsense, etc.

For their part, the more practical Wookiees believe that their new language successfully characterizes states of the world at the level of description required for the sort of successful action they in fact enjoy. But since it might be possible to do better, they take their commitment to

13. Since the language that evolves in this context is strictly in service of the payoffs of the game, if the game changes, then the evolved language should be expected to change. This provides a simple model for semantic drift in descriptive language over inquiry. It also provides a way of understanding Quine's distinction between *intuitive* and *theoretical* kinds (1969, 165). Here one might think of intuitive kinds as evolving in the context of one game and theoretical kinds in a different subsequent game, perhaps one with finer-grained payoffs. If the theoretical language is sufficiently successful, the theoretical kinds may replace intuitive kinds. But there need be nothing more than successful convention at work in the evolution of either language.

their current descriptions to be more a starting point for further empirical inquiry than a license for metaphysical speculation. It is unclear how their language will evolve in the future, but they hope that it will evolve to better satisfy stricter expressive demands that allow for further successful action. Insofar as they continue to impose stricter expressive demands, they may never find a perfectly stable reflective equilibrium between their descriptive language and how they update their signal-act dispositions; but in return for a degree of instability, they always allow for the chance of finding descriptive language better suited to the tasks at hand.

6. The Conventionality of Kinds. It is unclear the extent to which the evolution of a language in a syntactic game might faithfully model the evolution of our best descriptive languages. What a language evolves to track depends on the game in which the language evolves. Whatever the games may be in which our best descriptive languages have evolved, they are presumably more subtle than the signaling games considered here.

On the other hand, while coordinating states of the world with linguistic representations that in turn facilitate successful action may not be a complete characterization of scientific inquiry, it is arguably a fair description of a central aspect of scientific inquiry. And insofar as one finds this a compelling characterization, our best descriptive language may be expected to exhibit at least some of the characteristics of the simple languages that evolve in signaling games. In any case, the simple signaling games considered here show how it is possible for a perfectly successful language to evolve that tracks human dispositions and how the dispositions are updated rather than canonical metaphysical kinds.

On an appropriate occasion, a Wookiee may truthfully claim that the world is *blue*, *cold*, and *odd.* But the truth of such a claim does not depend on the structure of their language reflecting the structure of canonical kinds. Rather, it depends on the fact that their evolved language allows for the reliable individuation of states, up to the specificity required for successful action. That the terms of Wookieespeak do not track canonical natural kinds can be seen from the fact that there are no canonical kinds in the perfectly symmetric state space of the 8-state/2-term/3-sender game and the closely related fact that the Wookiees may have evolved an equally successful language that partitions the state space in an entirely incommensurate way.

While Wookiees may try to infer the structure of the world from their best descriptive language, their philosophical reflections may not yield at all what they want. What they can reliably infer from their successful kind language is that it is in reflective equilibrium with the dispositions that evolved it, and what they might have been tempted to take as natural kinds are purely conventional partitions of the state space contingent on

the available linguistic resources and the dispositional reinforcements of the game in which the language evolved.

One might tell a different evolutionary story starting with a state space that has canonical kinds baked into it, then consider conditions under which the resulting evolved language might in fact track the canonical kinds. But since a successful kind language may evolve when there are no canonical kind distinctions in the state space, the success of the kind language is not itself evidence that it tracks canonical kinds.

The natural metaphysician might object that the preoccupation here with the structure of language is misleading since it is only our best descriptive language *in the context of our best scientific theories* from which one is in fact justified in reading the metaphysics of the world. Quine, after all, held that the use of a natural-kind language is best understood as a sort of promissory note for a future theoretical account of its success. When one uses kinds in inductive inferences or to support counterfactual conditionals, "they may be seen perhaps as unredeemed notes; the theory that would clear up the unanalyzed underlying similarity notion in such cases is still to come" (Quine 1969, 169). The hope then might be that while our best descriptive language might exhibit conventional kinds contingent on the details of its evolution, our best theories will eventually allow us to distinguish between conventional kinds and genuine natural kinds in the descriptive language of the theory. More generally, one might hope that the co-evolution of successful theories and theoretical languages might better track genuine metaphysical distinctions than the evolution of a successful descriptive language alone. It turns out, however, that it is possible for a perfectly successful, but purely conventional, language to coevolve with an integrated and equally successful predictive theory. In order to see how this may occur, we will consider one last type of evolutionary game.[14]

A sender-predictor game is similar to a sender-receiver game except that after getting the signal from the sender, the receiver acts to predict the next state. A sender reporting "clouds" in the morning might, for example, produce a receiver predicting rain, thus carrying an umbrella in the afternoon. In the case of simple reinforcement learning, the sender's and receiver's dispositions to signal and to predict respectively are reinforced if the prediction is correct. The sender and receiver have coevolved *both* a successful descriptive language and a successful predictive theory if their dispositions evolve so that all of the receiver's predictions are

14. The idea of modifying sender-receiver games to include some sort of prediction came from a conversation with Michael Dickson. There is clearly much that might be done to develop such models beyond the simple deterministic-state model discussed here.

correct. If this happens, then one might think of both the language and the theory as coded for in the agents' dispositions to signal and to predict. In this case, it would be the language and theory together that are sufficient for successful coordinated predictive action.

The new degree of freedom in a sender-predictor game is the rule for the sequence of states. This rule represents the law of nature that the agents must learn in order to make successful coordinated predictions. Consider a 2-state/2-term sender-predictor game where the sequence of states is deterministic so that acting to predict state *B* when in state *A* is always correct and acting to predict state *A* when in state *B* is always correct. The sender's and receiver's actions are reinforced only if the receiver acts to predict *B* when *A* obtains and acts to predict *A* when *B* obtains.

Note that successful prediction in this particular deterministic 2-state/2-term sender-predictor game requires nothing more than an effective pairing of states and actions exactly analogous to the pairing in the 2-state/2-term sender-receiver game. Consequently, this sender-predictor game evolves in precisely the same way as the corresponding signaling game: the sender and receiver always approach a state where the sender's signal always leads the receiver to act to predict the correct next state of the world. As in the case of the corresponding sender-receiver game, the language that evolves in the sender-predictor game approaches reflective equilibrium with how the agents update their dispositions, but here the language and the integrated theory are in service of diachronic coordinated prediction.

While this game provides only a toy model of how scientific theories and descriptive language might coevolve, it shows how a conventional descriptive language might coevolve with an integrated predictive theory. This theory does not help one decide which kinds are conventional and which are canonical; rather, the theory relies on the evolved conventional partitions for its predictive success. Insofar as our best theories and descriptive languages may have coevolved similarly, one cannot count the theories to reliably distinguish between canonical and conventional kinds in the associated language.

While even our most successful theories and associated descriptive languages may not track canonical kinds, they do allow us to individuate states with the precision required for successful action. In this sense, carving nature at its canonical joints is unnecessary for successful descriptive inquiry.

REFERENCES

Barrett, Jeffrey A. (2006), "Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and the Evolution of Grammar", *UC Irvine Institute for*

Mathematical Behavioral Sciences Preprint (22 September 2006). http://www.imbs.uci.edu/tr/abs/2006/mbs06_09.

——— (2007), "The Evolution of Coding in Signaling Games", forthcoming in *Theory and Decision*.

Goodman, Nelson (1965), *Fact, Fiction, and Forecast*. New York: Bobbs-Merrill.

Herrnstein, Richard J. (1970), "On the Law of Effect", *Journal of the Experimental Analysis of Behavior* 13: 243–266.

Huttegger, Simon (2007a), "Evolution and the Explanation of Meaning", forthcoming in *Philosophy of Science.*

——— (2007b), "Evolutionary Explanations of Indicatives and Imperatives", *Erkenntnis* 66: 409–436.

Lewis, David (1969), *Convention* Cambridge, MA: Harvard University Press.

Quine, W. V. (1953), "Two Dogmas of Empiricism" in *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 20–46.

——— (1969), "Natural Kinds", in *Ontological Relativity and other Essays*. New York: Columbia University Press, 114–138.

——— (1981), "Five Milestones of Empiricism", in *Theories and Things*. Cambridge, MA: Harvard University Press, 67–72.

Roth, Al, and Ido Erev (1995), "Learning in Extensive Form Games: Experimental Data and Simple Dynamical Models in the Intermediate Term", *Games and Economic Behavior* 8: 164–212.

Skyrms, Brian (2004), *The Stag Hunt and the Evolution of Social Structure.* Cambridge: Cambridge University Press.

——— (2006), "Signals", Presidential Address, Philosophy of Science Association, PSA 2006.

Skyrms, Brian, and Robin Pemantle (2000), "A Dynamic Model of Social Network Formation", *Proceedings of the National Academy of Sciences in the USA* 97: 9340–9346.

Thorndike, Edward L. (1898), "Animal Intelligence: An Experimental Study of the Associative Processes in Animals" *Psychological Review Monograph Supplement* 2: 1–109.