Faithful description and the incommensurability of evolved languages

Jeffrey A. Barrett

Abstract   Skyrms–Lewis signaling games illustrate how meaningful language may evolve from initially meaningless random signals (Lewis, *Convention* 1969; Skyrms 2008). Here we will consider how incommensurable languages might evolve in the context of signaling games. We will also consider the types of incommensurability exhibited between evolved languages in such games. We will find that sequentially evolved languages may be strongly incommensurable while still allowing for increasingly faithful descriptions of the world.

Keywords   Skyrms–Lewis signaling games · Evolution of language · Incommensurability

1 Incommensurability of evolved languages

It has been imagined that the incommensurability of subsequent descriptive languages somehow counts against the associated theories providing increasingly faithful descriptions of the world. While such views are commonplace, it is perhaps appropriate to take Thomas Kuhn's formulation as paradigmatic. Kuhn noted that

> A scientific theory is usually felt to be better than its predecessors not only in the sense that it is a better instrument for discovering and solving puzzles but also because it is somehow a better representation of what nature is really like. One often hears that successive theories grow ever closer to, or approximate more and more closely to, the truth.

He then argued that the notion of a match between the descriptions of a theory and natural fact is incoherent (Kuhn 1996, p. 206). The argument is supposed to go

J. A. Barrett (✉)
Department of Logic and Philosophy of Science, UC Irvine, Irvine, CA 92697-5100, USA
e-mail: j.barrett@uci.edu

something like this. Since the descriptive languages associated with different paradigms are incommensurable, there can be no paradigm-independent way to describe what it would be for a statement to correspond to a fact. Hence our descriptions cannot be understood as increasingly faithful to the truth.

One of Kuhn's favorite examples of descriptive incommensurability was the relationship between descriptions in Newtonian mechanics and descriptions in Einsteinian special relativity. He noted that while the two languages share much of the same vocabulary, the meanings of their terms, at least as indicated by the truth conditions of the paradigm-specific statements that contain them, are radically different: "Newtonian mass is conserved; Einsteinian is convertible with energy. Only at low relative velocities may the two be measured in the same way, and even then they must not be conceived to be the same." (Kuhn 1996, p. 102) The reason that Einsteinian relativity cannot be understood as providing more faithful descriptions of the same physical world as described by Newtonian mechanics is that in abandoning the latter classical descriptions, "we have had to alter the fundamental structural elements of which the universe to which they apply is composed." (Kuhn 1996, p. 102). The incommensurability of subsequent descriptions then prevents us from conceiving of them as more or less faithful descriptions of the same world, which, in turn, rules out the view that empirical science provides an ever improving approximation to descriptive truth.[1]

I will argue that the incommensurability of descriptive languages in empirical inquiry is compatible with the view that each of the evolved descriptive languages allows for faithful descriptions of the world, and further, that the languages that evolve in the context of empirical inquiry should be expected to allow for increasingly faithful descriptions of the world. The argument here turns on understanding how incommensurable languages evolve in the context of Skyrms–Lewis signaling games (Lewis 1969; Skyrms 2008).

2 Skyrms–Lewis signaling games

There are more and less elaborate types of Skyrms–Lewis signaling games (Skyrms 2008; Barrett 2006, 2007, 2009). We will start with one of the simplest. A basic Skyrms–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 on each play of the game. The sender then observes the state and sends a term to the receiver, who cannot directly observe the state of the world. The receiver observes the term then chooses an act. Some acts count as successful given the state; otherwise, the act fails. On the two-player signaling game, the sender and receiver might learn from their successes and failures on repeated plays of the game. Whether they learn at all, what exactly they learn, and how quickly they learn it depend on how they update their conditional dispositions to signal on a state (for the sender) and to act on a signal

---

[1] See Bird (2008) for a reconstruction and criticism of Kuhnian arguments against convergent realism.
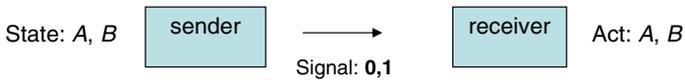
Fig. 1 A basic Skyrms–Lewis signaling game with one sender and one receiver

(for the receiver) in response to success and failure in matching the receiver's action to the state observed by the sender. If the sender's and receiver's conditional dispositions evolve so that the sender's signals lead the receiver to act in a way that is more successful than chance, then they have evolved a more or less efficient language, perfectly efficient if the evolved language always leads to successful action.

Consider a basic 2-state/2-term Skyrms–Lewis signaling game with simple reinforcement learning (see Fig. 1).[2] 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 this basic urn learning rule, there is no penalty for the act failing to match the state. The game is then repeated with a new state of the world.

This model represents act-based, rather than strategy-based, learning.[3] The number of balls of each type in the sender's and the receiver's urns represent their *conditional propensities*. The agents' *response functions* determine the conditional probability for a particular signal on a state or for a particular act on a signal given their current propensities. Here, and throughout the paper, we suppose that the response functions are proportional in that the conditional probability for a particular signal or act is the ratio of the associated conditional propensity for the signal or act to the total conditional propensity. The conditional probabilities, so determined, represent the *conditional dispositions* of the agents. And the learning

---

[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 model social network formation, and by Skyrms (2004, 2008) to model learning in the context of Lewis signaling games. Alternatively, one might consider such signaling games for a population of senders and a population of receivers or for a single population of agents who randomly exchange roles and track how the distribution of agent types evolve in the populations over repeated plays (see Huttegger 2007). Here, however, we will consider the evolution of conditional dispositions in the context of repeated plays between the same agents.

[3] Barrett and Zollman (2008) discuss this distinction and a description of the general structure of act- as opposed to strategy-based learning models.

rule describes how they update their conditional propensities on success and failure in action.

The simple reinforcement learning rule represented in urn learning has a long psychological pedigree. It is an instantiation of Richard Herrnstein's (1970) *matching law* account of choice, where the probability of choosing an action is proportional to the accumulated rewards. And Herrnstein's *matching law* was itself a quantification of Thorndike's *law of effect* (1898) for the conditioning of stimulus response relations by experience.

Argiento et al. (2009) have proven that perfectly efficient languages evolve with probability one in 2-state/2-term signaling games with evenly distributed states of nature and simple reinforcement learning. While the proof is nontrivial, 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 both evolve and learn the new meaningful language together. That they have done so is reflected in their new conditional dispositions and their subsequent track-record of successful action.

The situation is more complicated for signaling games with more (or fewer) states or terms or if the statistical distribution of states is biased (see Barrett 2006; Huttegger 2007). In such modified games, suboptimal equilibria may develop and prevent convergence to perfect signaling. Nevertheless, such systems typically evolve a language on even simple reinforcement learning that does better than chance in individuating the states of nature.

3 A two-sender signaling game and richer linguistic structure

More complicated signaling games allow for the evolution of languages where syntax plays a role in meaning. One of the simplest such games is a two-sender, one-receiver signaling game where there are four states of nature and four corresponding receiver acts but where each sender only has two terms that he might use to represent the state. The receiver knows which sender sent each signal (one sender, for example, always sends his signal first), but neither sender knows what the other sent. Here, since there are four states and four acts but each sender has only two terms, a perfectly expressive language can only evolve if the senders somehow learn to use the available syntactic structure (the order in which the terms are sent), to code for the four states. This requires each agent to evolve a partition of the world that is systematically related to the partition evolved by the other agent.

Consider a 4-state/2-term/2-sender signaling game where each of the two senders observe the complete state of the world then each sends a signal of either *0* or *1* to a single receiver (see Fig. 2). We will start again 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

Fig. 2 A signaling game with two senders and one receiver

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.

Computer simulations indicate that the agents here typically do evolve a perfectly efficient language that allows them to represent each of the four successful state-act pairs. Suboptimal 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.

On a successful run of the 4-state/2-term/2-sender signaling 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 "**A0B0**" representing state *0*, "**A0B1**" representing state *1*, "**A1B0**" representing state *2*, and "**A1B1**" representing state *3* (as indicated in Fig. 3).

Here each sender's signal selects a kind of state, and the specification of the two kinds together selects a single state at the level of specificity required for successful action given what is salient to the agents.

Such combination signaling has been observed in nature where the available basic terms are insufficient to capture the states of nature that matter for successful action of the organisms. In a recent study, Arnold and Zuberbühler (2008) have recorded free-ranging putty-nosed monkeys combining distinct alarm calls in order to convey different meanings. They have found that the call combinations are used to convey at least three types of information: the event witnessed by the male, the caller's identity, and whether he intends to travel. Further, the empirical evidence suggests that the meaningful combination signals are assembled in a rule-governed

Fig. 3 An evolved 2-sender kind language with **A**'s terms on the left and **B**'s above

A Simple Kind Language

way from more basic morphemes. Arnold and Zuberbühler conclude that it is the small vocal repertoires of the monkeys that may have favored the evolution of combinatorial signaling (Arnold and Zuberbühler 2008, p. R203).

4 Reinforcements and learning

Tracking how agents evolve a signaling language is subtle. The language is characterized by the evolved coordinated dispositions of the agents. It is a perfectly efficient language if the agents' evolved conditional dispositions to signal and act favor the same map from states to acts as their dispositions to update their dispositions to signal and act. The states, signals, and acts specified in the description of a signaling game are, in turn, simply part of the representation of the agents' dispositions to update their dispositions to signal and act. It is in this way that what in fact matters to the agents and how much it matters determines what it would be for an evolved signaling language to be successful, fixes the overall structure of the signaling game, and drives the evolutionary process.

Given their response functions, what matters to the agents and how much it matters is represented by their reinforcement functions. One specifies the reinforcement functions by identifying the states of nature, signals, and acts that matter to how the agents update their propensities then describing how their conditional propensities are updated given each particular combination of states, signals, and acts that may obtain. The agents' reinforcement functions provide a full specification of both what distinctions are relevant to their learning and precisely how they are relevant—less formally, they encode what is attended to by the agents and what they count as success.

More subtle reinforcement functions may lead to correspondingly more subtle languages and evolutionary histories, but the structure of the evolved language is not somehow smuggled into the model when one specifies the reinforcement functions of the agents. The reinforcement functions for real agents are not a matter of stipulation or convention; rather, they are a matter of dispositional fact that one either gets right or wrong in the model. Moreover, it is possible, as we will see, for incommensurable languages to evolve in the context of precisely the same reinforcements.

That said, the agents' reinforcement functions do constrain the possible structure of a successfully evolved language by determining the standard of success, setting the stage on which the evolution of propensities occurs, and driving the evolution of the propensities by refining them against the standard of success. A successfully evolved language then represents a reflective equilibrium where the agents using the language behave in such a way that they in fact reinforce exactly those conditional dispositions exhibited in their linguistic behavior.

5 Incommensurability of independently evolved languages

Incommensurable languages evolve most readily when the languages evolve independently and the reinforcement functions of the agents evolving each language

are different. Consider a land with two populations: a fraternity of crocs and a fraternity of zebras. Suppose that there are two types of state that matter to crocs (no zebras and zebras) and that they have evolved a primitive but successful term language in a way that is well-modeled by a 2-state/2-term/1-sender signaling game with reinforcement learning. There are consequently two possible croc signals ("**0**" and "**1**") corresponding to the two croc-salient states, and each signal leads to the appropriate action in crocs who hear it.

Suppose further that there are four types of state that matter to zebras (no croc, one croc, two crocs, and three crocs) and that they have independently evolved a primitive kind language in a way that is well-modeled by a 4-state/2-term/2-sender signaling game. There are consequently four possible signals ("**A0B0**", "**A0B1**", "**A1B0**", and "**A1B1**") corresponding to the four zebra-salient states, and each leads to appropriate action in zebras who hear it. Note that, for zebra action, "**A0**" means less than two crocs, "**A1**" means two or more crocs, "**B0**" means an even number of crocs, and "**B1**" means an odd number of crocs (as represented in Fig. 4).

That the crocs and zebras evolve different languages is a result of the differences in their reinforcements. The crocs only care whether or not zebras are present. They do not care how many zebras nor do they care what other creatures might be present, and they only distinguish between two possible terms. For their part, the zebras care about how many crocs there are up to three. They do not care about situations where there are more than three crocs nor do they care about whether or not other creatures are present. Like the crocs, the zebras only distinguish between two possible terms, but different combinations of the terms also matter to zebras. It is this last aspect of how zebras reinforce their propensities that allows them to evolve a kind language, a language with a very different structure than the term language of their croc neighbors.

While both languages are descriptive of the same world, Croc and Zebra are fundamentally incommensurable. There is no term-by-term translation between the languages. Nor is there any statement-by-statement translation between the languages. The different reinforcements of the two fraternities evolve languages with different expressive strength, different domains of applicability, and different fundamental structures.

One might imagine that, while there is sense in which Zebra and Croc are descriptive of the same world, from our perspective at least, their languages are only

Fig. 4 The kind language evolved by the Zebras

Zebra

	B0	B1
A0	0 crocs	1 croc
A1	2 crocs	3 crocs

incommensurable because they are descriptive of different aspects of the world. But incommensurable languages might also evolve to describe precisely the same aspects of the same world.

Suppose that the same four types of states that matter to zebras also matter to a fraternity of ducks, but that the ducks evolve a successful term language in a way that is well-modeled by a 4-state/4-term/1-sender signaling game. Here both the ducks and the zebras care about the same states of nature. The difference is that the ducks, unlike the zebras, have enough basic terms that they do not need to evolve combinatorial signaling. Consequently, the ducks evolve a term language where each of their four terms represents exactly one of the zebra-salient states. While both Duck and Zebra distinguish between the same states of nature, there is no term-by-term translation between the two languages; rather, terms in Duck correspond to combinations of terms in Zebra.

One might imagine that the evolution of incommensurable languages requires a difference in the reinforcement functions of the agents. This would cover the last example. While they care to distinguish the same states of nature, different linguistic resources are available to the ducks and zebras. But incommensurable languages might also evolve for agents with the same linguistic resources.

Suppose a fraternity of goats begins with precisely the same linguistic resources and reinforcements as zebras, and hence is well-modeled by a 4-state/2-term/2-sender signaling game, but evolves the kind language represented in Fig. 5.

While Goat and Zebra allow for the same individuation of states, they correspond to different evolved different partitions of the state space, so there is no term-by-term translation between the two languages. Here, for example, "**B0**" means the same thing in both languages, but no term in Zebra translates "**A0**" in Goat.

One might seek comfort in the fact that while the structure and the meanings of the terms are incommensurable, whole expressions in Duck, Zebra, and Goat are intertranslatable. But the sorts of incommensurability exhibited between evolved languages, even under identical conditions, can be significantly more subtle.

Consider the sort of incommensurability between signaling languages that occurs when, for one reason or another, they have evolved to be suboptimal relative to the states distinguished in the agents' reinforcement functions. As perhaps the simplest case, suppose that the fraternities of rats and pigs have precisely the same second-order conditional dispositions and that each fraternity evolves a term language in a

Fig. 5 The kind language evolved by the goats

Goat

	B0	B1
A0	2 crocs	1 croc
A1	0 crocs	3 crocs

a   **b**

State   Signal   Act   State   Signal   Act

1   1   p   1   1   1   s   1

                                1-p   1-s

2   2   2   3   2   3

          q   r

          1-q   1-r

3   3   3   2   3   2

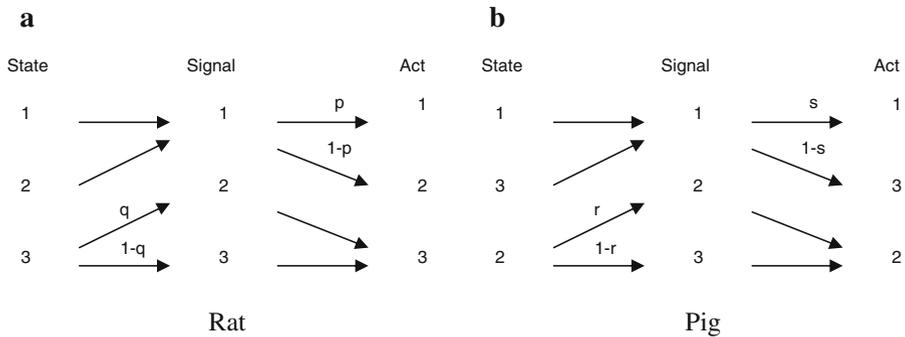Rat   Pig

Fig. 6 **a** and **b** represent the suboptimal languages evolved by the rats and pigs respectively. *Arrows* indicate the states that lead to signals and the signals that lead to acts

way that is well-modeled by a 3-state/3-term/1-sender signaling game with reinforcement learning, but suppose that the languages evolve to *different* partial pooling equilibria. More specifically, suppose that the rats evolve the language represented by Fig. 6a, where the sender uses signal **1** for both states 1 and 2 and randomizes between terms **2** and **3** in state 3 (with probabilities q and 1-q) and the receiver randomizes between act 1 and 2 when he sees signal **1** (with probabilities p and 1-p) and does act 3 when he sees signal **2** or **3**. Consequently, "**1**" means state 1 or 2 and "**2**" and "**3**" both mean state 3.

Similarly, in the same environment and with the same dispositions, the pigs evolve a language represented by Fig. 6b, where "**1**" means state 1 or 3 and "**2**" and "**3**" both mean state 2. So Rat distinguishes between states 1 and 3 but not 1 and 2, while Pig distinguishes between states 1 and 2 but not 1 and 3. The two languages are consequently neither term-by-term intertranslatable nor statement-by-statement intertranslatable.[4]

The upshot is that strongly incommensurable languages may evolve in the same world under precisely the same agent dispositions. But, while mutually incommensurable, each of the independently-evolved languages allows for faithful descriptions of the same world.

6 Refinement and revolution in sequentially evolved languages

Clearly, any sort of incommensurability that might obtain between independently evolved languages for different groups of agents might obtain between sequentially evolved languages for a fixed group of agents given enough time *and the evolution of appropriately different reinforcement functions*. One difference, however, between independently and sequentially evolved languages is that insofar as a fixed group of agents gradually evolve their linguistic dispositions, temporally

---

[4] This sort of incommensurability typically occurs for independently evolved languages whenever the linguistic resources are insufficient to represent every state that might matter to the agents. Since there is no perfect language here, different evolutions typically compromise differently.

proximal steps in the evolution will always correspond to at least roughly intertranslatable languages.

We will consider the evolution of subsequent languages in the context of a more subtle 4-state/2-term/2-sender signaling game. Rather than supposing that an agent's conditional dispositions are only reinforced when the act exactly matches the state, we will suppose that such reinforcements are determined by the degree of appropriateness of the act to the state. It might, for example, be ideal for a particular sort of agent to swim in water that is 100F in that conditional propensities are maximally reinforced when this happens. But it might also be acceptable to swim in water that is 80F, which might lead to somewhat more modest reinforcement. The agent might be ambivalent to 70F water in that it does not reinforce either way. And finding himself in 33F water might negatively reinforce the conditional propensities that put one there.

Suppose that the possible states of the world are numbered 0, 1, 4, and 5, that the possible actions are numbered 0, 1, 4, and 5, and that the senders' and the receiver's propensities are updated just as in the 4-state/2-term/2-sender signaling game that characterized zebra reinforcements earlier except that the magnitude of the change in the conditional propensities here is $m - r|s - a|$, where m is the maximum reinforcement, r is a constant scale factor, s is the state number, and a is the act number. This reinforcement function describes a maximum reinforcement of conditional propensities when the state and act match exactly ($s = a$). The constant r and the nonlinear assignment of proximity to states and acts represent how much it matters to the reinforcement of conditional propensities when the state does not match the act exactly. The nonlinear assignment of proximity means that ($s = 0$, $a = 1$) is less of a mistake than ($s = 1$, $a = 4$), even though the agents recognize no actions intermediate between $a = 1$ and $a = 4$. Note that negative reinforcements of the agents' conditional propensities are possible if $m < 5r$.[5]

Since the evolved language in a signaling game is represented in the relations between the positive conditional propensities of agents, allowing for *negative* reinforcements provides an effective way of evolving a new language by first unlearning the old language. On the other hand, if $m \ll 5r$, then the negative reinforcements of propensities may prevent the evolution of new positive propensities, and hence unlearn the first language but also prevent any new language from evolving.[6]

Now, consider a signaling game where $m = 2$ and $r = 1$ for the nonlinear reinforcement functions represented by $m - r|s - a|$. Just as with the zebras, a maximally efficient kind language often evolves, but the likelihood of converging to a suboptimal language here is greater than with simple reinforcement learning. This is because the nonlinear reinforcement functions that represent the agents' second-

---

[5] Throughout the remainder of this section, the qualitative descriptions of the behavior of the models is based on computer simulations. Unless otherwise indicated, the results described are for repeated runs of the models described in the text with 10^7 plays per run. For the two-stage models described later, where the reinforcement function changes between stages, each stage was run for 10^7 plays.

[6] Given the proportional response function we are using, negative-and zero-valued propensities do not represent dispositions, so if a punishment would push a given conditional propensity below 1.0, the propensity is set to 1.0.

order dispositions strongly distinguish between state 0 *or* 1 and state 4 *or* 5 but only weakly distinguishes between states 0 and 1 and between states 4 and 5—indeed, the agents are still rewarded if act 1 occurs, say, when state 0 obtains. Consequently, it often happens that the evolved language distinguishes between state 0 *or* 1 and state 4 *or* 5, but fails to distinguish between state 0 and state 1 or between state 4 and state 5. And when a language does evolve to capture the fine-grained distinctions between states, this typically occurs only after the language has first evolved to capture the coarse-grained distinctions.

This is perhaps just as one would expect. Languages should be expected to evolve most quickly to capture those distinctions that matter most to the agents, then only evolve more slowly to capture more subtle distinctions. *Changing* the reinforcement functions, however, can quickly sharpen a language that has only evolved to capture the coarse-grained distinctions so that it also captures the finer-grained distinctions. Moreover, appropriate changes in the reinforcement functions may allow an agent to escape a suboptimal equilibrium that they could not have escaped otherwise; though, of course, if the reinforcement functions are sufficiently different, what it means to be optimal will change.

Consider a signaling game just like the last one but with two stages.[7] The first-stage reinforcement functions are represented by $m - r|s - a|$, where $m = 2$ and $r = 1$ as before, but for the second stage the first-stage reinforcements are replaced by reinforcement functions where $m = 2$ and $r = 3$. While the first-stage reinforcements reward the sender and receiver if act 1 occurs in state 0 (though not as well as if the state matches the act), the revised reinforcements punish the agents for not making such fine-grained distinctions (though not as much as they are punished for not making the coarse-grained distinction). While the new reinforcement functions might be thought of as representing a change in the norms or aims of the agents, they, at root, just represent the new way that the agents update their conditional propensities—one that might have evolved from the first-stage reinforcement functions given a change in the environment that presents the agents with tasks that require a finer-grained individuation of states or that might represent higher aspirations as the agents become accustomed to their past success.[8]

In any case, the second-stage reinforcements here are capable of updating conditional propensities in a way not possible with the first-stage reinforcements. In particular, the agents may now unlearn suboptimal conventions by lowering the corresponding conditional propensities, which may quickly lead to a maximally efficient language that captures both the coarse- and fine-grained distinctions implicit in the reinforcement functions. Indeed, when the second-stage reinforcements do evolve a new language, it is typically just a conservative refinement of the

---

[7] Michael Dickson first suggested such two-stage evolutions as a way of investigating how the sender and receiver work together to evolve a successful language. The idea was to investigate whether changes in dispositions in the second change are symmetric between the two or whether one or the other is more likely to make the necessary accommodations for successful action in the new context.

[8] Understanding how reinforcement functions evolve is central to any account of how agents learn to learn. A model might start with a primitive learning strategy that is then used by the agents to evolve context-specific reinforcement functions, but exactly how this should work is a topic of considerable interest.

old where redundant linguistic resources are reallocated to capture the finer-grained distinctions.

Changing the reinforcement functions may, however, evolve a language that is strongly incommensurable with the language evolved in the first-stage. This happens, for example, if the agents are punished so severely by the new reinforcement functions for failing to capture finer-grained distinctions that their propensities are pushed low enough that they must retool and start over again in evolving the new language. Insofar as the punishments have erased the prior propensities, the new language here evolves independently of the old.

The likelihood of the new language being incommensurable with the old language and the degree to which is it incommensurable depends primarily on the similarity of the reinforcements that drive the evolutions. The reinforcement functions must change for a new language to evolve if the old evolution is near an equilibrium, but, as in the case of conservative refinement above, modestly incommensurable languages may evolve in the short run even if the new reinforcement functions are associated with the same individuation of states, signals, and acts and has the same pairing of best acts to states. More strongly incommensurable languages evolve if the new reinforcement functions identify best acts with states differently. And the strongest sorts of incommensurability evolve when an entirely new way of individuating states, terms, and acts is required to characterize the agents' second-order dispositions. In this case, the structure of the signaling game itself must be changed in order to faithfully represent the dispositions of the agents.[9]

As an example of a relatively modest but nontrivial sort of incommensurability, consider a two-stage signaling game where the first-stage reinforcement function is $m - r|s - a|$, where $m = 2$ and $r = 1$ and the second-stage reinforcement function is the same but with states 1 and 4 exchanged, and $m = 0.5$ and $r = 1$. Here the second-stage reinforcement function respects the same individuation of states as the first, it just associates best acts to states differently: the first-stage reinforcement function has act 1 as maximally reinforcing in state 1 and act 4 as maximally reinforcing in state 4 while the second-stage reinforcement function has act 1 as maximally reinforcing in state 4 and act 4 as maximally reinforcing in state 1. The second-stage reinforcement function also involves a punishment for not getting the fine-grained distinctions right.

When a perfectly efficient kind language evolves in the first stage, a different, but also perfectly efficient language, often evolves under the new reinforcements. On one run, for example, "**A0B0**", "**A0B1**", "**A1B0**", and "**A1B1**" might evolve to mean state 0, state 1, state 5, and state 4 respectively in the first stage, then might evolve to mean state 0, state 4, state 5, and state 1 in the second stage. These sequentially evolved languages exhibit the same sort of incommensurability as between Zebra and Goat. Here new **B0** has the same meaning as old **B0** and new **B1** has the same meaning as old **B1**. But there is no translation for new **A0** or new **A1** in the old language and no translation for old **A0** or old **A1** in the new language. So

---

[9] One might argue that in this case the game changes in such a way that the agents would never be able to describe fully the new game in their old language no matter how subtle is was. This is related to the possibility of a complete descriptive language discussed in the next section.

while the two languages allow for the same individuation of states, they are not term-by-term intertranslatable.

Insofar as one takes situations where there is such an exchange in optimal acts to be less likely than situations requiring refinement, one might expect the more radical sort of incommensurability that evolves here to be correspondingly more rare. But, that said, this and yet stronger sorts of incommensurability, like that between Rat and Pig, are exhibited in this model when, for example, neither the first nor second evolved language is optimal given the states and actions distinguished in the agents' reinforcement functions.

7 Incommensurability and the faithfulness of description

One might imagine that, since we can understand precisely how each of the evolved signaling languages discussed above represents states and actions, there can be no genuine incommensurability between them. The point here is akin to the complaint that the strong incommensurability that Kuhn imagines between Newtonian mechanics and Einsteinian relativity cannot be genuine since, in order to show just how different they are, Kuhn himself compares and contrasts how they describe the world.

A salient difference between Kuhn's story (however one might read him) and ours, however, is that we are considering weakly expressive primitive languages, not languages that are meant to represent the current limits of faithful description. Rather than directly argue for the incommensurability of our most expressive descriptive languages, the argument here is by analogy. The sense in which signaling languages may be incommensurable given their expressive resources is easily understood given the expressive strength of our best descriptive language; but, just as such primitive languages may be incommensurable given their expressive resources, our best descriptive languages may be incommensurable given their expressive resources.

One might imagine that we will eventually evolve a descriptively complete language that will provide intertranslatability between all languages that allow for any sort of faithful description of the world. Given the formal paradoxes that plague expressively strong languages, however, it may be incoherent to suppose that we will ever have such a language. But even if such a language were logically possible, there are good methodological reasons for never believing that we have it. Just as with primitive signaling languages, our best descriptive languages may be complete in the sense that they allow for the sort of individuation of states required for successful action given our current purposes but prove to be entirely inadequate for more demanding future purposes. Further, insofar as we are committed to empirical inquiry and insofar as empirical inquiry involves a commitment to search for ever more faithful description, to imagine that our current descriptive language is complete would be to place an arbitrary, and likely disastrous, constraint on future empirical inquiry.

That said, for smoothly evolving agent dispositions, one should expect sequentially evolved languages to be more or less intertranslatable in the short

run. And insofar as one takes what matters to empirical science to be stable, one should typically expect relatively modest sorts of incommensurability in the medium run, perhaps often akin to conservative refinement. Finally, even in those cases where it proves difficult to find a faithful translation between subsequent descriptive languages associated with different empirical theories, we are sometimes able to construct a backward-looking translation between the different modes of description in the context of our current best theories. Indeed, we are often driven to devise such constructions when we want our new theories to account for the explanatory successes of our old theories in a way that somehow preserves the old explanations.[10]

But, and this is the important point, whether or not and the extent to which subsequently evolved languages are incommensurable is in principle independent of whether or not and the extent to which they allow for faithful description of the world. This is illustrated by the fact that all of the signaling languages we have considered, while sometimes strongly incommensurable, allow for entirely faithful, if ultimately incomplete, descriptions of the same world.

A maximally efficient language is one that has evolved to individuate those states that are in fact relevant to the agents in updating their propensities. The evolution of increasingly expressive languages, then, depends on the evolution of increasingly demanding reinforcement functions that represent the evolved saliencies of the agents. Insofar as agents do evolve descriptive languages that satisfy increasingly demanding saliencies, their evolved languages will allow for increasingly faithful, though perhaps incommensurable, descriptions of the world.

8 Conclusion

Some degree of incommensurability should be expected between languages that evolve subject to different constraints, but we have also seen that incommensurable languages may evolve for agents with the same dispositions in the same environment. Even strongly incommensurable languages, however, may allow for perfectly faithful descriptions of the same world given what matters to the agents. Further, languages that allow for increasingly faithful descriptions of the world may be expected to evolve insofar as one expects that our future descriptions will answer to more demanding evolved saliencies. Indeed, insofar as we require ever richer predictions and explanations and insofar as we believe that the evolution of language is driven by such demands, we should expect our future descriptive languages, even when incommensurable, to allow for increasingly faithful descriptions of the world in precisely those senses we care about most.

---

[10] How we construct such translations between subsequent descriptive languages is discussed in Barrett (2008).

topics. I have benefited from Brian's striking combination of philosophical originality, precision of thought, and care for the truth at very turn.

References

Argiento, R., Pemantle, R., Skyrms, B., & Volkov, S. (2009). Learning to signal: Analysis of a microlevel reinforcement model. *Stochastic Processes and their Applications, 119*(2), 373–390.

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology, 18*(5), R202–R203.

Barrett, J. 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*, MBS06-09.

Barrett, J. A. (2007). Dynamic partitioning and the conventionality of kinds. *Philosophy of Science, 74*, 527–546.

Barrett, J. A. (2009). The evolution of coding in signaling games. *Theory and Decision, 67*(2), 223–237.

Barrett, J. A. (2008). Approximate truth and descriptive nesting. *Erkenntnis, 68*(2), 213–224.

Barrett, J. A., & Zollman, K. (2008). The role of forgetting in the evolution and learning of language, Manuscript.

Bird, A. (2008). Thomas Kuhn, *The stanford encyclopedia of philosophy*. In E. N. Zalta (Ed.). Accessed June 15, 2008, from http://plato.stanford.edu/entries/thomas-kuhn/.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243–266.

Huttegger, S. (2007). Evolution and the explanation of meaning. *Philosophy of Science, 74*, 1–27.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.

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

Roth, A., & Erev, I. (1995). Learning in extensive form games: Experimental data and simple dynamical models in the intermediate term. *Games and Economic Behavior, 8*, 164–212.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge: Cambridge University Press.

Skyrms, B. (2008). Signals. Presidential address, philosophy of science association, PSA 2006. *Philosophy of Science, 75*(5), 489–500.

Skyrms, B., & Pemantle, R. (2000). A dynamic model of social network formation. *Proceedings of the National Academy of Sciences in the USA, 97*, 9340–9346.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplement, 2*(4), 1–109.