ON THE COEVOLUTION OF THEORY AND LANGUAGE AND THE NATURE OF SUCCESSFUL INQUIRY

JEFFREY A. BARRETT

ABSTRACT. Insofar as empirical inquiry involves the coevolution of descriptive language and theoretical commitments, a satisfactory model of empirical knowledge should describe the coordinated evolution of both language and theory. But since we do not know what conceptual resources we might need to express our future theories or to provide our best future faithful descriptions of the world, we do not now know even what the space of future descriptive options might be. One strategy for addressing this shifting-resource problem is to track the predictive and linguistic dispositions of inquirers rather than their theories and conceptual resources. Sender-predictor games, a variant of Skyrms-Lewis sender-receiver games, provide very simple models for the coordinated coevolution of predictive and linguistic dispositions. Such models explain *how it is possible* for (1) predictive and descriptive dispositions of inquirers to coevolve, (2) term-wise incommensurable, but nevertheless descriptively faithful languages, to sequentially evolve, and (3) a sort of underdetermination to occur where inquirers might satisfy their descriptive and predictive aims by revising their linguistic dispositions, their theoretical dispositions, or both. Such models also provide an elementary characterization of what it might mean for descriptions of the world to be faithful and for empirical inquiry to be successful. In doing so they provide a relatively weak, but perfectly clear, endogenous account of epistemic norms.

1. KNOWLEDGE AND THE EVOLUTION OF DESCRIPTIVE LANGUAGE

On the standard account of propositional knowledge, an agent X knows proposition P *if and only if* (1) X believes P, (2) P is true, and (3) X has rational justification for believing P. In contrast, belief-revision models of knowledge seek to describe how one's local descriptive, predictive, and explanatory commitments do and should evolve as one gains new evidence. On a Bayesian belief-revision model, for example, one fixes a descriptive language, sets prior probabilities over a fixed set of hypotheses expressed in the language, then considers how one's coherent degrees of belief evolve as one conditions on new evidence.

Date: June 17, 2011.

By considering how one's beliefs evolve given one's local commitments more generally, belief-revision models capture the local, diachronic nature of empirical inquiry and avoid the embarrassment of requiring that one somehow justify all of one's beliefs and commitments at once. They also allow for degrees of local epistemic commitment that covary with degrees of local evidentiary support, rather than requiring, as is typical on the standard account, a special type or degree of rational justification for a belief to count as genuine knowledge.

While belief-revision models have salient advantages over the standard account for modeling the evolution of knowledge, even a cursory consideration of actual empirical inquiry indicates that both the set of hypotheses under consideration and the descriptive language used to express these hypotheses change over time. In particular, empirical inquiry nearly always involves descriptive *discard*, *invention*, and *drift*. For example, in turn, reference to *luminiferous aether* is not required to account for the propagation of light on our best physical theories, there is no classical analogue to a quantum-mechanical system in a *superposition* of classical properties, and, while very different notions in the context of classical mechanics, there is a sense in which *mass* and *energy* may be taken to mean the same thing in the context special relativity. A satisfactory account of empirical inquiry then must address the *shifting-resource problem*: insofar as we expect our theories and descriptive language to coevolve, we currently lack the descriptive resources that would be required even to characterize our future theoretical options.

The shifting-resource problem is a problem for both the standard account and for belief-revision models of knowledge. In what sense might one have rational justification for believing any theoretical description of the world on the standard account of knowledge if one holds that one likely does not yet have a language entirely suitable for expressing descriptive truth? How should one update one's degrees of belief if one holds that the descriptive language one is using to express possible beliefs is itself changing over time? Moreover, it is unclear even what the space of options should be for revising one's descriptive language insofar as this is itself a matter for future empirical inquiry.[1]

---

[1]The shifting-resource problem is closely related to the problem of how to handle the catch-all hypothesis in a Bayesian model of inquiry, which is, in turn, closely related to what Kyle Stanford calls the problem of unconceived alternatives[17]. The shifting-resource problem is also related to the problems posed for rational inquiry by the threat of diachronic Kuhnian incommensurability[14]. The fundamental problem is that, insofar as we expect our descriptive resources to evolve in inquiry, we currently lack the resources even to express our best future theoretical options. Stanford takes this to pose a problem for scientific realism. The suggestion

Rather than try to characterize the space of possible future conceptual options, one might endeavor to track inquiry by considering how inquirers' descriptive, predictive, and explanatory *dispositions* coevolve. But even if one accepts the suitability of tracking conceptual resources by tracking the dispositions of inquirers, there remains a serious problem here of feasibility. Providing a fine-grained account of the evolution of the dispositions of real inquirers would be extraordinarily difficult, perhaps simply impossible. That said, one might model the coevolution of such interrelated dispositions for very simple, highly idealized inquirers. The thought is that such simple models might provide an elementary understanding of *how it is possible* for descriptive and theoretical resources to coevolve. And, along the way, one might gain some degree of insight into how best to understand epistemic norms in the context of such dispositional models. Such an account will be as thin as such models are over-simplified, but insofar as the account is clear, it might nevertheless represent significant progress in our understanding of inquiry and its norms.

This paper is concerned with showing how descriptive and predictive dispositions might coevolve in the context of sender-predictor games with bounded reinforcement learning with punishment. Such games allow one to explain *how it is possible* for (1) the descriptive and predictive dispositions of inquirers to coevolve with relatively few resources, (2) term-wise incommensurable, but nevertheless descriptively faithful languages, to sequentially evolve, and (3) a sort of underdetermination to occur where inquirers might satisfy their descriptive and predictive aims by revising their linguistic dispositions, theoretical dispositions, or both. Such models also provide a very basic characterization of what it might mean for descriptions of the world to be considered faithful and empirical inquiry to be understood as successful. In doing so they provide an endogenous account of epistemic norms. While this account is very simple and relatively weak, it is at least perfectly clear.

2. Sender-predictor games

Sender-predictor games provide elementary models for the coevolution of descriptive and predictive dispositions. Sender-predictor games are variants of Skyrms-Lewis sender-receiver games.[2]

---

here is that the shifting-resource problem is a problem for anyone who wishes to provide a concrete model of inquiry insofar as such a model may require one to characterize future conceptual options.

[2]For discussions of standard Skyrms-Lewis sender-receiver games and variants see [12], [16], [9], [1], [8], [7], [5], and [15]). The idea of a sender-predictor game came from conversations with Michael Dickson. Dickson has subsequently studied sender-predictor games that are significantly more subtle than those described here. Of particular interest are models where the laws of nature are not deterministic. In such models, assuming a simple learning dynamics for the agents that

In the most basic standard Skyrms-Lewis *sender-receiver game* there are two agents: a sender and a receiver. The sender observes the state of nature, then sends an signal. The receiver observes the signal then performs an action that either matches the state of nature and is successful or does not match the state of nature and is unsuccessful. If the act is successful, then the disposition that led to each agent's last action is reinforced; otherwise, it may be weakened. Indeed, a successful action here is simply an action that generates a result that in fact leads to the reinforcement of those dispositions that produced the action (insofar as the learning dynamics allows for these dispositions to be further reinforced), and an unsuccessful action is an action that generates a result that does not lead to the reinforcement of the dispositions that led to the action (and, insofar as the learning dynamics allows, may lead to the weakening of these dispositions).

On such a game, the sender's signals are initially meaningless, but as her dispositions to signal conditional on the state of nature and the receiver's dispositions to act conditional on the signal evolve, the sender's signals become meaningful insofar as they serve as the basis for successful coordinated action. Whether a perfectly successful signaling system evolves in such a game depends on the number and distribution of states of nature, the agents' signaling resources, and the precise learning dynamics one considers.[3]

Similarly, in the most basic *sender-predictor game*, the sender observes the prior state of nature, then sends a signal. On receiving the signal, the receiver then performs a predictive action that is either successful or unsuccessful depending on the posterior state of nature at a later time. The sender might, for example, observe the color of the morning sunrise, then clap once or twice. On hearing the signal, the receiver might either prepare for a picnic outside or a lunch indoors for their afternoon meal. If this predictive act is successful, that is, say, if the inquirers end up either picnicking outside in the sun or lunching indoors warm and dry while watching the rain through a window, then the disposition that led to the sender's signal and the disposition that led to the receiver's action is reinforced; otherwise these dispositions might be weakened. Since only the sender sees the morning sunrise and since the weather in the afternoon in part determines the success of

---

closely resembles the dynamics described below, the agents are typically able to successfully evolve a descriptive language and successfully predict the *most likely* evolution of nature. We expect to describe these models in a future co-written paper.

[3]See [12] for the original description of signaling games in the context of traditional game theory. See [15], [1], [13], [5], and [9] for descriptions of the basic game and variations in the context of evolutionary game theory and for discussions of conditions under which successful signaling might evolve.

the receiver's predictive action, the sender and receiver must coevolve dispositions that are reliable both in signaling prior states of nature and in predicting posterior states of nature in order to be successful. But, again, that an action is described as successful here means nothing more than that the dispositions of the inquirers that lead to the action are in fact reinforced given the actual result of the action.[4]

In discussing such models it is useful to distinguish between the inquirers' first- and second-order dispositions. The sender's and receiver's dispositions to signal and to act may be thought of as their *first-order dispositions*. In contrast, their *second-order dispositions* are their dispositions to update their first-order dispositions on the basis of the actual results of their descriptive and predictive actions. The inquirers' second-order dispositions, then, determine how they learn from experience. The suggestion will be that they may also be taken to provide an endogenous notion of what it means for an action to be successful in the context of the model. But we will return to this point later.

A slightly more complicated sender-predictor game is represented in Figure 1 below. Here there are four prior states of nature, four signals the sender might use, and four possible predictive actions for the receiver. We will suppose that the inquirers learn using bounded reinforcement with punishment. On this sort of learning dynamics, the inquirers' first-order dispositions may be thought of as represented by the proportional contents of urns and a description of how the contents of the urns determine their actions, and their second-order dispositions may be thought of as represented by how the contents of the urns are updated given their experience.

More specifically, regarding the inquirers' first-order dispositions, the sender has one urn for each possible state of nature labeled $0, 1, 2,$ and 3. Each of her urns begins with one ball of each type of signal she might send labeled $A, B, C,$ and D. The receiver has one urn for each type of signal. Each of his urns begins with one

[4]If the dynamics of nature is deterministic, the sender-predictor game will be formally equivalent to a sender-receiver game with the same overall structure. That the same game can be interpreted two very different ways, as a game involving the evolution of language for successful coordinated action or as a game involving the evolution of language and successful prediction, is evidence for how closely intertwined the evolution of language and theory may be. That the same model might be plausibly interpreted multiple ways indicates that the domains of the different interpretations may in fact exhibit the common structure represented by the model. This line of argument is a recurring theme in this paper. Note, however, that if the dynamics in nature is not deterministic, the sender-predictor game has no standard sender-receiver counterpart (unless one considers, for example, sender-receiver games with noise or where the right action is randomly determined as might be the case if the sender and receiver were cooperating to play mixed strategies in a game against other agents).

ball of each type of predictive action he might take labeled $0, 1, 2,$ and 3. The sender observes the prior state of nature, draws a ball at random out of the corresponding urn, with each ball equally likely, then sends the corresponding signal. The receiver observes the signal, draws a ball at random out of the urn corresponding to the signal, with each ball equally likely, then performs whatever predictive action is indicated on the ball drawn.

To stipulate that the inquirers learn by bounded reinforcement with punishment is to endow them with second-order dispositions of a particular sort. More specifically, if the receiver's predictive action was *successful*, then the sender and receiver put the ball they drew back in the urn they drew it from and add to that urn another ball of the same type if there are fewer than $N_{\max}$ balls of that type already in the urn. If the predictive action was *unsuccessful*, then they each put back the ball they drew if and only if it was the last ball of its type, which will typically weaken the associated dispositions. Given these second-order dispositions, there will always be at least one ball of each type in each urn, and never more than $N_{\max}$ balls of any particular type.

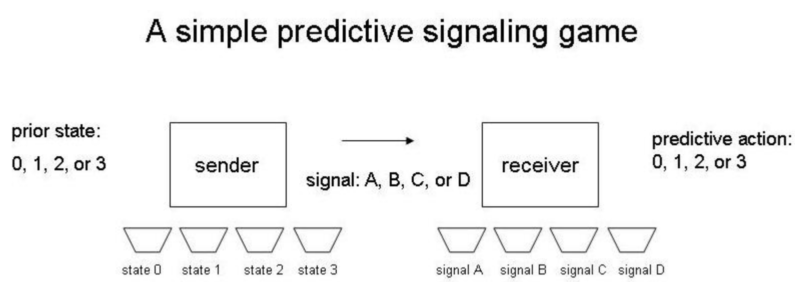

Figure 1. A sender-predictor game.

In this model the sender and receiver begin by randomly signaling and randomly predicting. Any initial success they have is consequently the result of blind luck. As their first-order dispositions evolve under the influence of their second-order dispositions, however, the sender and receiver are found on simulation nearly always to evolve a set of nearly optimally successful, and systematically interrelated, descriptive and predictive dispositions.[5]

[5]On this model, for $N_{\max} = 1000$ and 1000 runs each with 10^6 plays, for example, the sender and receiver nearly always (0.993) evolve a set of nearly optimal (0.994) dispositions. The exact details of the learning dynamics do not matter much as similar results are obtained for a wide range of parameter values for this sort of bounded reinforcement learning with punishment. What matters for evolving successful descriptive and predictive dispositions here is that (i) there be a maximum number of each type of ball in each urn (so bad habits do not get too strongly ingrained),

Since only the sender observes the prior state of nature, the receiver acts for both agents, and the success of the receiver's action is in part determined by the posterior state of nature, the inquirers must evolve both dispositions to faithfully represent the prior states of nature and dispositions to reliably predict posterior states in order to be successful. Put another way, to be successful, the inquirers must evolve a language sufficient to represent the states of nature, use this language to truthfully describe nature, and learn the patterns of nature with sufficient precision to take advantage of these patterns for the sake of predictive action. Whenever they do so, their first-order dispositions evolve in such a way that they are stable given their second-order dispositions. Their first-order dispositions then code for a sort of practical knowledge that allows the inquirers to satisfy the norms represented by their second-order dispositions.

In addition to showing how it is possible for very simple inquirers to evolve the ability to faithful describe and predict states of nature, this model also shows how closely intertwined these dispositions may be. Indeed, they are so closely intertwined in this particular model that the shared ability of the inquirers to reliably signal and predict might be described equally well as (1) the sender telling the receiver what the prior state of nature is then the receiver predicting the posterior state or (2) the sender predicting the posterior state then telling the receiver what predictive action to take. The inquirers would have to be playing a more subtle game in order to distinguish between these two readings. The following two games arguably allow for just such a distinction.[6]

There are a number of natural generalizations of such basic sender-predictor games. Consider, for example, a game involving two senders, each of which fails to have the resources to fully specify the prior state of nature as a signal to the receiver. Each sender observes the prior state of nature, then sends a signal as

(ii) no type of signal or predictive action ever go to extinction in any urn (so there is always a possible escape from suboptimal dispositions), and (iii) both reinforcement and punishment, or weakening, of first-order disquisitions be possible (since without reinforcement, the inquirers would be unable to learn; and without punishment they would be unable to forget suboptimal dispositions and hence lack the means of escape). On such a learning dynamics it is possible for the inquirers' first-order dispositions to randomly wander away from success, but they will quickly return and spend most of their time almost ideally successful in their actions. See [10] for a discussion on the role of forgetting in evolutionary games.

[6]A yet richer game, in the relevant sense, would be one where the same term is sometimes used to promote present action and sometimes used to promote future action. If the two uses of the same term are assumed to be synonymous, such a game would allow one to determine who is describing prior states and who is predicting future states in a sharper sense than the games below.

before. Each sender has one urn for each possible prior state of nature labeled $0, 1, 2,$ and 3. Each of their urns begin with one ball of each type of signal each sender might send labeled 0 and 1. The receiver has one urn for each type of *composite signal* the two senders might send labeled $00, 01, 10,$ and 11. Here the first term is the term sent by sender A and the second by sender B. Each of the receiver's urns begins with one ball of each type of predictive action he might take labeled $0, 1, 2,$ and 3. The second-order learning dispositions of the inquirers are given by bounded reinforcement with punishment as described earlier.

As one might expect with a model this simple, there are a number of natural interpretations. One would be to think of the signals sent by the two observers as information from different sources that go into a single predictive law. The receiver takes a predictive action that is contingent on each of the signals and is either successful or unsuccessful given the inquirers' second-order dispositions. If the act is successful, then the first-order disposition that led to each inquirer's last act is reinforced; otherwise it is weakened. Note that there is no mechanism here for the senders to coordinate their signals directly. The only information they get regarding the other sender's behavior comes from their joint success and failure given the receiver's actions.

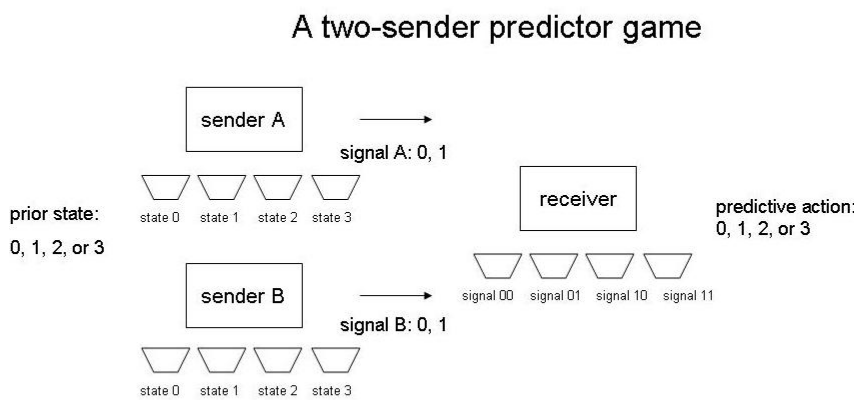

FIGURE 2. A more subtle sender-predictor game.

Here, again, the senders and receiver start off randomly signaling and randomly acting, but, as they learn from experience, they typically evolve a set of nearly optimal, systematically interrelated, linguistic and predictive dispositions.[7] More

[7]For parameters akin to those discussed in the last game, the senders and receiver typically (0.995) evolve a set of nearly optimal (0.994) linguistic and predictive dispositions. Again, such results hold for a wide range of parameter values on bounded reinforcement learning with punishment.

specifically, the senders typically evolve coordinated partitions over the prior states of nature that correspond to something like the extension of natural kind terms, and the receiver nearly always makes successful predictive actions given their signals.[8]

In contrast with the last game, it is less natural here to understand the senders as simply telling the receiver what predictive action to take since, in order to be successful, what the receiver does must depend on the signal he gets from *each* sender. Since neither signal is sufficient to determine the receiver's predictive action, it is arguably the receiver who is making the predictions.

Now consider what happens in this last game when the manner in which first-order dispositions are reinforced changes after a successful set of descriptive and predictive dispositions have initially evolved. Again, there are a number of ways to interpret such a game. One might understand the changing reinforcements as changing patterns in nature or as the changing norms of the inquirers. The question at hand, however, is simply whether the modeled inquirers will be able to evolve *new* dispositions that answer to different natural laws or different norms. If so, one might understand the inquirers as having successfully carried out something akin to a mini-Kuhnian revolution.[14]

On the sort of learning dynamics we are considering here, the modeled inquirers typically do in fact carry out just such a mini-revolution. Moreover, while their new first-order dispositions are often strongly *incommensurable* with their initial descriptive dispositions, their descriptions of the world may be perfectly faithful in both their old and new languages.

The game is the same as the last except that the inquirers change how they update their first-order dispositions after they have initially evolved a set of successful descriptive and predictive dispositions. More precisely, in the first stage of this game, the inquirers evolve successful descriptive and predictive dispositions just as described in the last game; then, in the second stage of this game, the inquirers begin to reinforce posterior action 2 on prior state 1 and posterior action 1 on prior state 2 and weaken the dispositions associated with any other actions involving these states. Again, this change in how the agents reinforce may reflect either changes in the regularities of nature or changes in the agents' norms, or both.

As before, the inquirers typically evolve successful descriptive and predictive dispositions in the first stage of the game; more to the point here, in the second stage of the game, they typically also adapt to the new reinforcements and evolve

[8]The language that evolves in this game is similar to the simple compositional language used by putty-nosed monkeys. See [2] and [3]. For a discussion of the evolution of natural kind terms see [7].

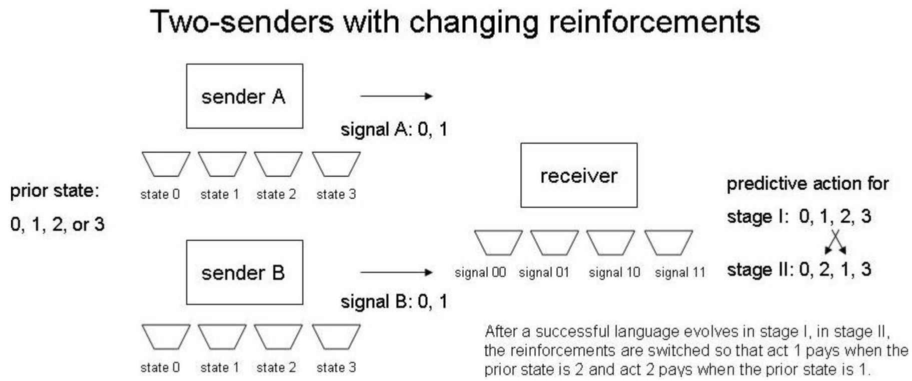

FIGURE 3. A sender-predictor game with changing second-order dispositions.

a second set of nearly optimally successful descriptive and predictive dispositions.[9] But it is underdetermined precisely how the modeled inquirers will evolve the new set of dispositions. The *receiver* may evolve to make predictions differently or, less frequently, the two *senders* may evolve new coordinated partitions that classify the states of nature differently.[10]

When the senders evolve a new descriptive language, their old and new languages are often incommensurable in the sense that they are not term-by-term intertranslatable. The term-wise incommensurability of their languages does not, however, represent any failure in their ability to faithfully describe the world they inhabit. While the model of inquiry here is simple and idealized, this point nevertheless entails a quite general philosophical moral: that a particular model of inquiry may lead to incommensurable ways of describing the world cannot in itself be taken as providing any evidence whatsoever for the skeptical conclusion that inquiry on that model is not tracking descriptive truth. If they were unable to evolve the ability to faithfully represent prior states and to reliably predict posterior states, the inquirers could not ultimately evolve to be uniformly successful in action. Of course, this does not mean that the inquirers have the resources to characterize the precise senses in which their descriptions of the world are and are not faithful.

[9]For parameters akin to those described in the last two games, the inquirers here typically (0.972) evolve a second set of nearly optimal (0.993) descriptive and predictive dispositions that answer to the new regularities of nature and/or changed second-order dispositions.

[10]In simulations of the game as described here, the *receiver* evolved to make predictions differently with probability 0.581 and the *senders* evolved new coordinated partitions that classify the states of nature differently with probability 0.419. This lack of symmetry is interesting. Here it may be the result of the fact that it is harder for the senders to evolve a new coordinated language than for the receiver to evolve a new way of making predictions.

In order to better understand the sort of term-wise incommensurability exhibited here and how it might evolve, consider two concrete cases from two simulated runs of this game. In Case I, as represented in Figure 4, it is the receiver who has evolved the new dispositions. The senders still associate the same terms with the same prior states of nature, but the receiver has changed how he makes predictive actions on the senders' signals.

Case I: receiver evolves new dispositions to act

	prior state	A sends	B sends	receiver does	
old language and theory	0	1	0	0	stage I
	1	0	1	1	
	2	1	1	2	
	3	0	0	3	↓
new language and theory	0	1	0	0	stage II
	1	0	1	2	
	2	1	1	1	
	3	0	0	3	

FIGURE 4. Sometimes the receiver evolves new dispositions.

In Case II, as represented in Figure 5, it is the two senders who evolved the new dispositions. The senders associate different terms with the prior states of nature, but the receiver makes predictive actions precisely the same way as he did before the change in reinforcements. On an interpretation of the model where the senders are describing the prior states of nature on their signals, their initial and subsequent languages are strongly term-wise incommensurable. Here, as indicated in Figure 6, B_1 means the same thing in both languages, but no term in the old language translates the new term A_0.[11]

[11]The sort of termwise incommensurability exhibited here is relative tame. Since the agents can, in full statements, individuate precisely the same states of nature, while they lack the ability to translate terms between languages, they can translate statements. More subtle types of incommensurability, however, may evolve in such games. See, for example, [5] for a discussion of the sort of incommensurability that may be exhibited between languages corresponding to different partial pooling equilibria. In this case, there is neither term-wise not statement-wise commensurability between the evolved languages. Further, while the agents cannot individuate precisely the same states in such a case, their descriptions of the world may nevertheless be equally faithful insofar as they may allow for equally faithful overall precision in description.

Case II: senders evolve new dispositions to signal

	prior state	A sends	B sends	receiver does	
	0	0	1	0	
old language and theory	1	1	0	1	stage I
	2	0	0	2	
	3	1	1	3	↓
	0	0	1	0	
new language and theory	1	0	0	2	stage II
	2	1	0	1	
	3	1	1	3	

FIGURE 5. Less often the senders evolve new dispositions.

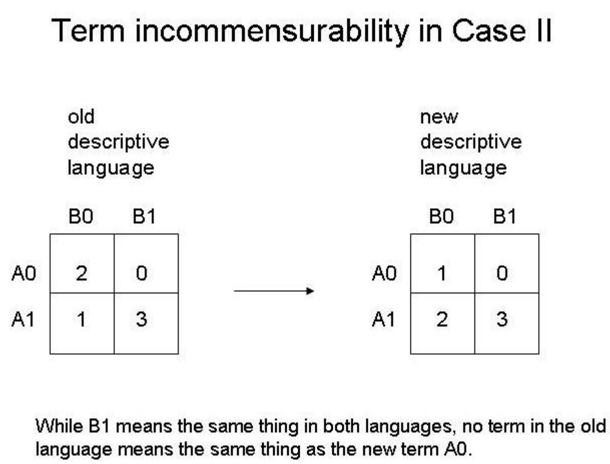

FIGURE 6. Incommensurable descriptive languages when the senders evolve new dispositions.

That either the senders or the receiver might revise their dispositions in order to answer to the new reinforcements illustrates again how closely intertwined descriptive and predictive dispositions may be on such models and how each sort of disposition is necessarily involved in successful action. Insofar as such models show how it is possible for equally faithful but incommensurable language to evolve in turn, they also show why even the possibility of strong incommensurability in description is not in itself evidence against our ability to faithfully describe nature. While successful inquiry here consists in nothing more or less than the establishment of an appropriate harmony between nature and the second-order dispositions of the inquirers, this harmony cannot be established without the inquirers in fact evolving the ability to faithfully represent prior states of nature and reliably predict posterior states of nature.

3. An Endogenous Notion of Epistemic Norms

Sender-predictor games allow one to explain *how it is possible* for (1) descriptive and predictive dispositions to coevolve (and to coevolve in such a way as to accommodate changes in nature or in the inquirers' second-order dispositions), (2) term-wise incommensurable, but nevertheless descriptively faithful languages, to evolve in turn and (3) a sort of underdetermination to occur where the inquirers might satisfy their descriptive aims by revising their descriptive or their theoretical dispositions or both. Such models also provide a weak but fully endogenous and perfectly clear account of epistemic norms.

As suggested, *successful inquiry* on such models may be understood as evolution toward a reflective equilibrium between the inquirers' first-order dispositions to signal and predict, their second-order dispositions to reinforce their dispositions to signal and predict, and the actual results of their signaling and predictive actions given the nature of the world they inhabit. More specifically, the inquirers have been successful in inquiry when a reflective equilibrium has evolved between their second-order dispositions and their descriptive and predictive actions such that their second-order dispositions to change their first-order dispositions no longer in fact change how they describe or predict. Similarly, *error* can be understood as occurring whenever the inquirers' first-order dispositions to generate signals and predictions produce an event that leads them to revise their current first-order dispositions given their second-order dispositions. Using such colorful language, then, successful inquiry might be said to consist here in the elimination of descriptive error for the purpose of successful action.[12]

While there need be nothing normative in any absolute sense in the establishment of such a reflective equilibrium, at least from the local perspective of the inquirers, there is a sense in which they are getting precisely what they want from inquiry for the purpose of successful action when the results of their descriptive signals and predictive actions are in harmony with their second-order dispositions. On this

[12]The pragmatic account of successful inquiry suggested here fits well with, for example, C.S. Peirce's characterization of scientific inquiry. See "Some Consequences of Our Four Incapacities (1868)," "The Fixation of Belief (1877)" and "How to Make our Ideas Clear (1878)", chapters 3, 7, and 8 of [11], respectively. Peirce's theory of signs also allowed for the coevolution of language and theory. Indeed, sender-predictor games might be taken as providing a very simple dispositional model for his late theory of signs where the interpretant in his account is coded for in the dispositions of the inquirers. See [4] for an introduction to Peirce's semiotics.

view, what counts as successful inquiry is in part determined by the nature of the inquirers, but it is also a function of the nature of the world the inquirers inhabit. When the inquirers reach a reflective equilibrium between their first-order dispositions, their second-order dispositions to update their first-order dispositions, and what actually happens when they describe and predict, they are getting what they want. But they are only getting what they want when their first-order dispositions evolve in such a way that their second-order dispositions are satisfied by what actually happens when they describe the world and act on these descriptions, which, in turn, requires that the inquirers do in fact evolve the ability to faithfully describe the nature of the world and successfully predict, at least to the degree and precision required for successful action given their second-order dispositions, in order to be successful even by the own lights. If the inquirers' second-order dispositions were to become more demanding, correspondingly finer-grained descriptions of the world and more reliable predictions in terms of these more exacting descriptions would be required for success by their lights.[13]

All of this is just to draw attention to the fact that success in inquiry here cannot be taken as in any way being *constituted* by the nature or state of the inquirers alone. If the modeled inquirers are successful and we change the nature of the world they inhabit, they are typically no longer successful, even by their own lights, at least not until they evolve a new set of first-order dispositions that coordinates the properties of their world with what they want as represented by their second-order dispositions. This is not a deep or subtle point. The evolved first-order dispositions of the inquirers must track *both* the structure of the payoffs *and* the nature of the world they inhabit in order for inquiry to be successful. So, while there is a sense in which the inquirers' evolved dispositions are entirely in service of their second-order dispositional norms, their actions cannot satisfy these norms without their signals in fact evolving to faithfully represent nature and their predictions in fact evolving to capture the dynamics of the states of nature. Of course, the precise sense in which their descriptions may or may not be faithful need not be something that the inquirers themselves can even express in their language as it stands at any given time. Indeed, insofar as the inquirers believe that their best future descriptions of the world will require descriptive resources that they do not yet possess, they will also believe that they cannot characterize how their current best descriptions fail.[14] Rather, they exhibit the faithfulness of their descriptions and the reliability of their

[13]See [6] for a discussion descriptive nesting in inquiry.

[14]See [6] for an extended discussion of this point.

predictions in the successful actions they make that are based on these descriptions and predictions.

Successful inquiry in the sense exhibited by sender-predictor games simply requires that the inquirers' first-order dispositions evolve in such a way that their second-order dispositions are well coordinated with what in fact happens when they act. While this is a relatively weak account of epistemic norms, it does require that the inquirers in fact evolve the ability to faithfully describe and reliably predict states of the world they inhabit. And it may not differ in kind from yet-to-be-described stronger accounts of epistemic norms. The thought is that stronger notions may be found, just as the present relatively weak notion was found, in the context of more detailed and subtle evolutionary models of inquiry, perhaps models that include the evolution of explanatory dispositions.

4. Conclusion

The shifting-resource problem requires that we understand knowledge in a way that allows for the coevolution of descriptive language and theory. Sender-predictor games allow one to explain how it is possible for descriptive and theoretical dispositions to coevolve to meet the shifting demands of inquiry without requiring an *a priori* or canonical space of linguistic or conceptual options. They also allow one to explain how incommensurable, but similarly faithful descriptions of nature may evolve in turn. Finally, such models provide a basic characterization of what it might mean to provide faithful descriptions of the world and reliable predictions, and, hence, what it might mean for empirical inquiry to be successful. In doing so, they provide a relatively weak, but perfectly clear and fully endogenous, account of epistemic norms. One might expect to find richer notions of normative inquiry in richer evolutionary models. A natural place to look would be in the context of evolutionary models that involve descriptive, predictive, *and* explanatory actions, where the inquirers' explanatory actions make use of their evolved descriptive resources.[15] While such models will invariably be more subtle and sophisticated than those discussed here, one should expect that a similar pragmatic notion of successful inquiry will be available there as well.[16]

[15]Tracking such explanatory dispositions would require a model that is significantly more subtle than anything considered in this paper. The evolution of explanatory dispositions might, for example, involve the evolution of dispositions to send signals that characterize the relationships between one's own first- and second-order dispositions to signal and predict and that may instill similar relations in the dispositions of other inquirers.

[16]I would like to thank Brian Skyrms, Michael Dickson, Elliott Wagner, Simon Huttegger, Tucker Lentz, and the participants at the Stanford Workshop for helpful discussions. I would also

References

[1] Argiento, Raffaele, Robin Pemantle, Brian Skyrms and Stas Volkov "Learning to Signal: Analysis of a Micro-Level Reinforcement Model," *Stochastic processes and their Applications* 119(2):373–390.

[2] Arnold, K., and Zuberbhler, K. (2006). "The Alarm-Calling System of Adult Male Putty-Nosed Monkeys, Cercopithecus Nictitans Martini," *Animal Behaviour*, 72, 643-653.

[3] Arnold, Kate and Klaus Zuberbhler (2008) "Meaningful Call Combinations in a Non-Human Primate," *Current Biology* 18(5): R202–3.

[4] Atkin, Albert (2010) "Peirce's Theory of Signs" *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/peirce-semiotics/. Accessed 14 March 2011.

[5] Barrett, Jeffrey A. (2009) "Faithful Description and the Incommensurability of Evolved Languages," *Philosophical Studies* 147(1): 123–137.

[6] Barrett, Jeffrey A. (2008) "Approximate Truth and Descriptive Nesting," *Erkenntnis* 68(2): 213–224.

[7] Barrett, Jeffrey A. (2007a) "Dynamic Partitioning and the Conventionality of Kinds," *Philosophy of Science* 74:527–546.

[8] Barrett, Jeffrey A. (2007b) "The Evolution of Coding in Signaling Games," *Theory and Decision* 67(2): 223–237.

[9] 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* MBS06-09.

[10] Barrett, Jeffrey A and Kevin Zollman (2009) "The Role of Forgetting in the Evolution and Learning of Language," *Journal of Experimental and Theoretical Artificial Intelligence* 21(4): 293-309.

[11] Houser, Nathan and Christian Kloesel (eds.) (1992) *The Essential Peirce: Selected Philosophical Writings Volume 1 (1867-1893)*. Bloomington and Indianapolous: Indiana University Press.

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

[13] Huttegger, Simon (2007) "Evolution and the Explanation of Meaning," *Philosophy of Science* 74: 1–27.

[14] Kuhn, T. S. (1996) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

[15] Skyrms, Brian (2010) *Signals Evolution, Learning, & Information*, New York: Oxford University Press.

[16] Skyrms, Brian (2006) "Signals," *Philosophy of Science* 75(5): 489–500.

[17] Stanford, Kyle (2006) *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.

UC Irvine, Irvine, CA 92697, USA
E-mail address: j.barrett@uci.edu

like to thank Jim Weatherall and Cailin O'Connor for helpful comments on an earlier draft of this paper.