On Distinguishing the Meaningless from the Meaningful: An Evolutionary Game Theoretic Approach to Ruth Millikan's Teleosemantics

Lindley Slipetz

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Arts
In
Philosophy

Lydia K. Patton
Joseph C. Pitt
Benjamin Jantzen

April 2nd, 2013
Blacksburg, Virginia

Keywords: Ruth Millikan, game theory, meaning, teleosemantics

On Distinguishing the Meaningless from the Meaningful: An Evolutionary Game Theoretic Approach to Ruth Millikan's Teleosemantics

Lindley Slipetz

ABSTRACT

What distinguishes a meaningless utterance from a meaningful term? While one might say that, within the context of Ruth Millikan's teleosemantics, it is a term's having a proper function that distinguishes it from a meaningless utterance, I propose that the distinction can be made with reference to the history of the term. Using evolutionary game theory, I offer a way to clarify the distinction between the meaningless and the meaningful. I reject the possibility of correlating meaning with an evolutionarily stable strategy as this does not seem to be consistent with how communication works or with Millikan's theory. Instead, when a term has meaning, the function category of that term corresponds to an evolutionarily stable state composed of both speaker and hearer strategies.

Table of Contents

1. Introduction

What distinguishes a meaningless utterance from a meaningful term? The purpose of this paper is to address this question in the context of Ruth Millikan's teleosemantic theory of language. In doing so, I will propose a sketch of a model of Millikan's definition of meaning development using evolutionary game theory. While I recognize the diversity of positions that have been presented as attempts to define meaning, I do not aim to provide an exhaustive account of meaning. Instead, I concentrate on the historical account of meaning as it has been presented by Millikan. The clarification of Millikan's position will help to offer support to other naturalistic theories of language, as well as to work in naturalized epistemology.

Millikan's theory of language serves as a basis for her naturalized epistemology. She writes that one motivation for the naturalized approach to philosophy is to "reconcile...our view of man as a natural creature and a product of evolution"[1] with our theories of language, thought, and knowledge. By viewing humans as products of evolution, we can conceivably see our capacities for language and thought to be evolutionary products as well. While, in this paper, I concentrate on how the meaning of a term can have an evolutionary, historical basis, such an approach is also compatible with evolution-based theories of the origin of language generally.

Numerous naturalized accounts of the origin of language use evolutionary game theory as a way to model how members of a community come to communicate and understand each other. For example, using David Lewis' sender-receiver games, Brian Skyrms has argued that more sophisticated forms of language can evolve like basic biological signals.[2] Just as signals transmit between neurons, we can think of the origin of a language as growing from such signal transmissions into coordinated information transfer or communication. Because of the proposed

---

1 Ruth Garrett Millikan, *Language, Thought, and Other Biological Categories* (Cambridge: MIT Press, 1984), 7.
2 See Brian Skyrms, *Signals: Evolution, Learning, and Information* (New York: Oxford University Press, 2010)

analogy between the origin of language and evolutionary biology, we can think of the history of language and meaning as following an evolutionary model. Thus, in this paper I aim to clarify Ruth Millikan's view in terms of evolutionary game theory. While I am remaining agnostic regarding the correctness of Millikan's position, I believe that exploring its compatibility with evolutionary game theory is a fruitful exercise for the development of naturalistic philosophy more generally.[3]

According to Millikan, the meaning of a term is its function, and this function explains the term's history of successful proliferation through a population. It is through this "stabilizing" function that a term acquires meaning. Although this condition seems to account for a clear distinction between meaningless utterances and meaningful terms, clarification is needed. Specifically, Millikan claims that when a term has meaning, the use of that term is an evolutionarily stable solution. This reference to evolutionary game theory will not do the work that Millikan requires of it without further elaboration on whether this solution is a strategy or a state.

The difference between evolutionarily stable strategies and states lies in what we require to remain stable given the introduction of mutant or novel strategies. "A strategy is a policy that an individual follows that determines how it will behave,"[4] and a state is a configuration of strategies. For instance, let *x* and *y* be strategies, series of rules for how an individual will act. Strategy *x* is an evolutionarily stable strategy when other strategies, like strategy *y*, cannot successfully invade a population in which every member uses *x*. Now, instead, let *z* be a state

---

[3] Evolutionary game theory has other applications throughout philosophy and elsewhere. It has the potential to model any evolving social or decision situation. As such, its applicability is no longer restricted solely to biological scenarios. Evolutionary games have been used in economic models, social contract and moral theories, cultural evolution and numerous other topics. Even here, though this is a biologically-based theory, the use of evolutionary game theory is meant to characterize language evolution.

[4] Elliott Sober, *Philosophy of Biology*, 2nd ed. (Boulder: Westview, 2000), 139.

with proportions of both the *x* and *y* strategies. State *z* is an evolutionarily stable state when there are no other state configurations that would result in the population having a higher fitness.[5] The importance of this distinction in the context of Millikan's theory is that it helps to determine the scope and diversity of what has evolved within the language community. For strategies, that a term currently has meaning would depend on the previous actions of individuals in a population evolving towards one strategy; and, for states, meaning occurs because of the resulting stability of the interactions between groups with different strategies in a population.

In this paper, I attempt to clarify the distinction between the meaningless and the meaningful as it appears in Millikan's work and argue that Millikan's evolutionarily stable solution is meant to be an evolutionarily stable state with mixed strategies. To begin, I will offer a more detailed picture of Millikan's view with an emphasis on history's role in establishing meaning. Second, I will introduce a puzzle that arises when attempting to distinguish the meaningless from the meaningful. Third, I will discuss Millikan's use of evolutionary game theory in order to offer a more clear account of when a term has meaning. Finally, I will conclude by proposing aspects of the evolutionary game theoretic frame that must be clarified in order to complete the model.

2. On Proper Functions and History

For Millikan, all meaningful terms have proper functions [6], which are a "matter of having been 'designed to' or of being 'supposed to' (impersonal) perform a certain function."[7, 8]

---

5 See section 3 of this paper for a more detailed analysis of this distinction.

6 Millikan stipulates that "actual body organs and systems, actual actions and purposive behaviors, artifacts...and many customs, etc., all have proper functions" ("In Defense of Proper Functions" 293). I recognize that drawing an analogy between such a diversity of objects and actions is controversial; however, as I stated at the beginning of this paper, I will be abstaining from discussing the plausibility of such comparisons. I am merely granting Millikan's account of language as one possibility for a history-based theory of language and using this as a starting point for the evolutionary game theoretic analysis that will be presented in section 3.

7 Millikan, *Language, Thought, and Other Biological Categories,* 17.

8 Although Millikan is explicitly using "design" to help define functions, it is important to note that this notion

Generally, "the meaning of a...language device[9] is its focused stabilizing function,"[10] which is dependent upon the hearer's (or consumer's) rather than the speaker's (or producer's) interpretation of the term. In preparation for understanding the use of history in distinguishing the meaningless from the meaningful, let us consider this view in more detail.

The proper function of a language device is determined by its history, not by its "present properties or dispositions."[11] While the successful history of the device may be biological in nature, we can also think of the device's history as simply analogous to an evolutionary history. Millikan writes that "products of evolution have in common with various other kinds of products the fact that they are reproduced[12] or continue to be proliferated because they, *rather than other*

---

should not be thought of as contrary to evolution (i.e., as being compatible with Intelligent Design). To see why this is the case, consider Millikan's distinction between the function of tools and the function of language. A physical tool, like a hammer, has been designed by someone to serve a particular purpose; and the form of the tool is fitting to serve this function. This is not the way we should think about the design of a language device: "Natural language devices are not (at least literally) devices once 'designed by someone' to serve certain functions" (*Language, Thought, and Other Biological Categories* 2). Unlike the physical tool, the design of the language device is the product of evolution. The success of a given device and its design is explained by its history rather than some original purpose. I am in no way positing an original purpose or any kind of design analogous to Intelligent Design.

9 By language device, Millikan means "words, surface syntactic forms, tonal inflections, stress patterns, punctuations, and any significant surface elements that a natural spoken or written language may contain" (*Language, Thought, and Other Biological Categories* 3). Notice that this definition of language device entails that certain expressions that might be considered meaningless within the context of other theories of language have meaning for Millikan. For example, certain noises or onomatopoeic expressions that have the proper history will be meaningful for Millikan. I recognize that this may be an unsatisfactory result for some; however, as the goal of this paper is to clarify Millikan's view, it is consistent with the current aim. Throughout this paper, I will use "language device" in this way. I consider terms to be a subset of language devices. As such, there will be some interchangeability between device, term, etc. I ask the reader to consider those definitions and constraints that apply generally to language devices to apply specifically to terms, words, etc. as well.

10 Millikan, *Language, Thought, and Other Biological Categories,* 52.

11 Millikan, "In Defense of Proper Functions," 288.

12 There is a bit more that can be said about the idea of "reproduction." Millikan makes a distinction between first-order and higher-order reproductively established families. A token member of a first-order reproductively established family is a direct copy of a type from the family, while a token member of a higher-order reproductively established family is similar to the type. Using the human heart as an example, the heart is not a member of a first-order reproductively established family because it is not a direct copy: my heart is not an exact reproduction of either my mother's or my father's heart. While my heart is not a member of a first-order reproductively established family, "it was produced under Normal conditions in accordance with the proper functions of certain of my genes which *were* directly copied from my parents' *genes*" (*Language, Thought, and Other Biological Categories* 25). Hence, in the use of a heart and similar biological examples, the heart is a member of a higher-order reproductively established family and the genes underlying the formation of the heart are members of a first-order reproductively established family.

things, have been associated with certain functions."[13] Let us look at Millikan's definition of direct proper functions as a way to clarify this idea.[14]

According to Millikan, a device with a direct proper function exists because of previous successful performances of that function.[15] More formally, "a function *F* is a direct proper function of *x* if *x* exists having a character *C* because by having *C* it *can* perform *F*," and *x*'s ability to perform *F* can be explained by the fact that there "'were things that performed *F* in the past due to having *C*.'"[16] This explanation for *x*'s ability to perform *F* is grounded in the history of *x* and serves to offer an account for why a given device has persisted. If *x*'s having *C* had not been positively correlated with the occurrence of *F*, then we would lack a plausible account for why *F* has maintained as opposed to *F′*.

Consider a common screwdriver as an example of this. In Millikan's view, our current screwdrivers exist with the functions that they have because previous screwdrivers successfully fulfilled the function of screwing in screws. Thus, screwdrivers have screwing in screws as a direct proper function. We could conceive of using the handle of a screwdriver to pound a nail into a thin wall. This is not a proper function because it does not explain the existence of our current screwdrivers. It was not because of a screwdrivers ability to pound nails that we now have screwdrivers. Because of the screwdriver's ability to fulfill its function, we now have screwdrivers that we can view as having proper functions.

---

[13] Millikan, *Language, Thought, and Other Biological Categories*, 27.

[14] Since this paper is focused on the role of history for proper functions generally, I will not say much in regards to the other main type of function, a derived proper function. We can think of derived proper functions in the following way. A language device has meaning in virtue of falling into specific patterns and relations; that is, the device is not a direct copy or tokening of previous devices. The prior successes of a device with a derived proper function that fit these general patterns act as a historically-grounded explanation for the continued standard use of these forms. For instance, a bee dances to indicate to other bees the location of nectar. These dances aren't direct copies of previous dances, but they fit a general form that allows other bees to correctly interpret the dance. For a more detailed account of derived proper functions, see Chapter Two of Millikan's *Language, Thought, and Other Biological Categories.*

[15] Millikan, "In Defense of Proper Functions," 288.

[16] Millikan, *Language, Thought, and Other Biological Categories,* 26.

Millikan claims that this is also how we should think about meaning and language. Our meaningful terms exist as they do now because they successfully performed some function in the past. Had some other expression better fulfilled this function, that expression would currently be in use rather than the other. For example, since the term "dog" successfully fulfilled its purpose of alerting other language users to the presence of a quadrupedal mammal, "dog," rather than some other utterance, now has meaning. Generally, the device's past successes help to explain why it, specifically, is reproduced and determines its current proliferation within a language community.

The stabilizing function of a device also figures into an explanation of meaning. It is that function that "tends at the same time to keep speakers using the device in standard ways and to keep hearers responding to it in standard ways."[17] Again, we can use the example of "dog" to understand this function. Consider the term, "dog," not having a stabilizing function. First, imagine that speakers use "dog" in totally arbitrary ways. One person might use it to express anger and another might use it as a greeting. With such random uses of the term, hearers would have no way to satisfactorily respond. They could greet the speaker of "dog," but they risk angering the already angry "dog" speaker. Alternatively, they could hide from who they think is an angry "dog" speaker, but actually confuse a greeting "dog" speaker. Because there is no standard use of the term, the hearers cannot respond in a standard way. In this scenario, the history of the term cannot explain its current use. While, on occasion, speaker and hearer uses may correspond, this would be merely coincidental. Hence, without standard speaker use, an utterance lacks a proper function.

Conversely, imagine if hearers react to "dog" in completely unpredictable ways. Some cheer, some stare blankly, and some run away. Because there is no standard response, a speaker

[17] *Ibid.,* 32.

has no expectation for how someone will respond. It seems that such an utterance would have no chance of gaining meaning. For Millikan's view, we require a historically grounded explanation for why one utterance rather than some other fulfills a specific function. If "dog" had no standard, correct responses, then some other term that was successful at causing hearers to respond correctly to certain quadrupedal mammals would be more fit to fulfill this function; and speakers would use this latter term. Thus, when a term has meaning, it has a standardizing function that depends on the correct hearer interpretation and speaker use.

Before continuing, I must stress a point that Millikan makes about standard and stabilizing functions:

> ...such a function should not be thought of as an invariant function or as an average function but as a function performed in a critical mass of cases of actual use, forming a sort of center of gravity to which wayward speakers and hearers tend to return after departures.[18]

Clearly an adequate account of language should not require that words are invariably used correctly. We have all been in situations in which a word or phrase is used in jest. Additionally, we have all been witness to at least one moment in which a speaker misuses a word. Thus, we are aware of occasions that support the idea that language use is not always correct. Also, from experience, it is clear that a term need not be used correctly on average for it to have meaning. Consider an analogous example. Imagine that one unfortunate day 90% of the world's screwdrivers broke and no longer successfully screwed in screws. According to Millikan, it is not the case that the proper function of screwdrivers would no longer be to screw in screws. Instead, we would simply have a lot of broken screwdrivers that no longer fulfill their proper function. Granting that we agree with Millikan on the analogy between the function of tools and language, language that is misused on average must not be said to lack a function.

---

[18] *Ibid.*

The final aspect of a meaningful language device is its focused function. A focused function is the culmination of a series of functions working towards one ultimate outcome.[19] For example, pushing the power button on a computer sets off all sorts of functions, all aiming towards the end goal of the computer turning on.[20] In the context of language, that a language device depends on other language devices for meaning need not detract from the former device having meaning itself. Consider the word "by" in the sentence, "I left my book by the window." Contrast this with "by" in "This essay was written by me." The other words in each sentence have functions on which each instance of "by" depends, resulting in "by" having two different functions. Still, this does not mean that the instances of "by" lack functions in themselves. In the former sentence, "by" functions as a preposition indicating proximity; and, in the latter, it indicates authorship. The focused functions of these language devices serve to give them meaning in context.

To summarize Millikan's view, the meaning of a language device is its stabilizing focused function; that is, it is the ultimate function of a serial chain of functions that work to keep speakers using it and hearers reacting to it in standard ways. The proper function of a term is that function that historically explains the current existence of the term as meaningful. Because the device successfully fulfilled its function in the past, it now has meaning. When we say that a language device has meaning, in Millikanian terms, we mean that the term has a historical precedent for correct use, that this stabilizing function explains the continued proliferation of

---

[19] *Ibid.,* 36.

[20] Here, one might be concerned that the power button of a computer has numerous focused functions. In addition to turning the computer on, the power button turns the computer off and can potentially put the computer into sleep mode. Millikan contends that she is not forced to be committed to a device only having one focused function (see *Language, Thought, and Other Biological Categories* 37). I agree with her on this point. If the computer is properly functioning, then the focused function of the power button of a turned off computer is to turn the computer on. On the other hand, if the properly functioning computer is on, then the focused function of the power button is to turn off. In both cases, the context in which the function is carried out is important in assessing the proper functioning of the device.

standard speaker uses and hearer responses.

In the next section, I will discuss the lack of clarity in differentiating meaningful devices from meaningless utterances; but, first, it is worth pausing to elaborate on the role of history. As was previously mentioned, in order for a term to have meaning, it must have a proper function; thus, there must be a historically-grounded explanation for its successful proliferation. For direct proper functions, this means that a language device has meaning due to being a direct tokening of a previous language device; and the existence and proliferation of a current token is explained by a history of successful use of the previous language device and other direct language ancestors.

It is important to note that the history requirement need not be read as a strictly temporal constraint. It is not the case that after *x* number of days, the language device has meaning. Instead we could think of the history requirement as saying that after *x* number of occurrences, the device has meaning. While these occurrences did happen in the past, the emphasis should be on the number of successful cases of use rather than on some temporal measure of time passed since the introduction of the device.

Still, I do not mean to suggest that we should expect there to be some definite number that determines when a function has been acquired. For example, it would seem rather arbitrary to say that after ten successful uses of an utterance, it suddenly has meaning. If this were the case then terms that took more repetition to establish stability, even just eleven occurrences, would be said to lack meaning. This seems like a needlessly strong constraint on what should count as meaningful. Later, I will take a different approach to describing a device's meaning; but, now, we turn to a potential problem.

3. The Meaningless Versus the Meaningful

A question arises from Millikan's account of meaning: what allows us to distinguish meaningful terms from meaningless utterances? It is clear that a term has meaning when it has a stabilizing focused function; however, it remains to be seen how we can use this to draw a distinction between the meaningful and the meaningless. Of course, one need not expect there to be a strictly defined line between the two; however, it will be an issue if Millikan's theory cannot provide a clear difference. In the next section, I will sketch out a history-based potential solution; but, for now, I will elaborate on this problem.

One might not see this as a legitimate concern: a meaningful term has a proper function, while a meaningless utterance does not. The meaningful device has this function determined by its history of successful direct reproductions. On the other hand, a meaningless utterance may have been successfully used on an occasion or two, but it has not had enough instances of success to acquire a proper function. Thus, the division between meaningless and meaningful is a matter of function.

If one simply relies on function to differentiate between the meaningful and the meaningless, the difference is seemingly circular. A device has meaning in virtue of its function. Hence, to say that what distinguishes something that is meaningful from something that is not by appealing to the latter's lack of function is to say nothing more than that a meaningful term has meaning and a meaningless utterance does not. Perhaps one will be satisfied with this answer, but I believe that Millikan's theory has more to offer. We must emphasize the role that history plays in determining the functions of meaningful devices: what distinguishes the meaningless from the meaningful is a difference in histories, and these histories can be defined using evolutionary game theory.

4. Evolutionarily Stable Solutions and Meaningful Terms

Hence, the question at hand is how are we to differentiate between the meaningless and the meaningful using an account of a device's history of successful proliferation. While we may not require that there be some concrete number of occurrences that indicates the difference between meaningless and meaningful, we could offer a more general description of how the history of a device determines its current status as meaningful. What I mean to say is that our goal should be to aim for a concept that captures Millikan's notion of history as it relates to meaningful devices.

Millikan seems to make an attempt at clarifying this role of history, writing that "once a specific language device in in place, using its conventional[21] forms in conventional ways is what evolutionary biologists call an evolutionarily stable solution."[22] This use of evolutionary game theory will help to provide us a model that does not depend on the specific number of occurrences of success: it is a way to more generally quantitatively assess the trends in a device's history that have resulted in it having meaning. The evolutionarily stable solution can serve as an explanation as to why this device rather than that utterance now stably functions within a language community.

Even so, Millikan's use of an evolutionarily stable solution does not provide us with enough information to clarify a device's status as meaningful. Evolutionarily stable solutions can refer either to strategies of an individual or states within a given population. In this section, I will analyze these two options in order to draw out the compatibilities and consequences of each approach. This clarification will serve as a sketch for how to build an adequate model of meaning acquisition for evolution-based theories of language.

---

[21] Although Millikan is referencing coordination problems in a later part of this passage, note that she is not using "convention" in a way similar to Lewis. For Millikan, a convention just is a device with a proper function.

[22] Ruth Garrett Millikan, *Language: A Biological Model* (Oxford: Clarendon Press, 2005), 191.

Following from Millikan's statement that the use of a language device is an evolutionarily stable solution, it seems plausible that she could mean an evolutionarily stable strategy. To understand evolutionarily stable strategies, let us turn to the paradigm case from John Maynard Smith and G.R. Price's "The Logic of Animal Conflict." In this work, Maynard Smith and Price provide an individual selection-based explanation as to why animals tend to use a limited war strategy rather than a total war strategy when in conflict.[23] By utilizing evolutionary game theory, they seek to use the results of their games as explanations for why one rather than some other strategy has persisted. This strategy, the evolutionarily stable strategy, is defined as "a strategy such that, if most of the members of a population adopt it, there is no 'mutant' strategy that would give higher fitness."[24] Additionally, it is considered a stable equilibrium because the introduction of alternative strategies will not alter the proportion of the population using the evolutionarily stable strategy. For the sake of the task at hand, I will give a brief overview of the game proposed by Maynard Smith and Price.

Consider a model of inter-species conflict between five strategies: mouse, hawk, bully, retaliator and prober-retaliator.[25] Each individual can act in three ways. First, they can use conventional tactics, C, those tactics that have less serious consequences for the victim. Second, they can use dangerous tactics, D, those tactics that are most likely to cause serious injury.

---

[23] For the curious reader, allow me to elaborate on some terminology. First, let us distinguish between total versus limited war strategies. A total war strategy is a kind of winner-take-all approach in which one uses the most destructive weapons and fighting styles in order to reap the most reward. A limited war strategy involves "inefficient weapons or ritualized tactics that seldom cause serious injury to either contestant" (Maynard Smith and Price 15). One might expect that males within the same species would adopt total war strategies in order to increase the likelihood of passing on their genes (less competition would mean more mating opportunities). Instead, it is found that animals tend to adopt the limited war strategy. There can be group- or individual-based explanations for adopting such strategies. An explanation based on group selection assumes that an adaptation was selected for because it was beneficial for the group. In the context of animal conflict, one would claim that limited war strategies are adopted because a total war strategy would be harmful to the species as a whole. Maynard Smith and Price deny this explanation because it seems unable to account for the number and diversity of species that adopt the limited war strategy. Instead, they claim that it is the benefit to individuals within a population that explains the prevalence of this strategy.

[24] John Maynard Smith and George Price, "The Logic of Animal Conflict," *Nature*, 146 (1973): 15.

[25] Detailed descriptions of each of the strategies can be found in Maynard Smith and Price on page 16.

Third, they can retreat, R. The five strategies are defined by rules that determine which moves will be used. For example, Maynard Smith and Price define the hawk strategy as a total war strategy: the hawk "always play[s] D [and c]ontinues the contest until he is seriously injured or his opponent retreats."[26] While, for the sake of brevity, I will not go into the details of all the strategies, it will suffice to say that the mouse, the retaliator and the prober-retaliator are limited war strategies. Next, probabilities are assigned to dictate when certain moves will be played and how much damage will be caused. Finally, a pay-off is defined for the end of each competition that measures "the contribution the contest has made to the reproductive success of the individual."[27] This determination of fitness depends on the advantage of winning and the disadvantages of injury, and wasting energy and time.[28]

In order to find the evolutionarily stable strategy, Maynard Smith and Price tested each strategy in relation to all others. The evolutionarily stable strategy was determined to be the retaliator strategy, the strategy for which no other strategies yielded a greater pay-off than the strategy against itself. The retaliator strategy against the retaliator strategy had a pay-off of 29.0, which is equal to or greater than the pay-offs of each of the other strategies against the retaliator. For example, the hawk strategy against the retaliator strategy had a pay-off of -18.1 to the hawk.[29] As Maynard Smith and Price write, "for [*x*] to be an ESS [an evolutionarily stable strategy], is it necessary that it be the most profitable strategy of a population almost entirely of [*x*]."[30] Because the evolutionarily stable strategy is the most profitable strategy, the population is

---

[26] *Ibid.,* 16.
[27] *Ibid.,* 15.
[28] *Ibid.,*15-16.
[29] Bully against retaliator yielded a pay-off of 11.9 to the bully. Prober-retaliator against the retaliator yielded a pay-off of 26.9 to the prober-retaliator. Mouse against the retaliator yielded a pay-off of 29.0 to the mouse. While the mouse strategy did equally well to the retaliator strategy, the definition of an evolutionarily stable strategy only requires that no other strategies yield a higher pay-off than the strategy against itself (Maynard Smith and Price 16).
[30] *Ibid.,* 16.

not vulnerable to the introduction of a new or mutant strategy; that is, the new strategy would not skew the population away from the evolutionarily stable strategy. More strongly, when there is an evolutionarily stable strategy, the entire population should adopt this strategy. While such an approach seems reasonable in the context of animal conflict, it remains to be seen how this can apply to meaningful devices.

When trying to frame a term with meaning as an evolutionarily stable strategy, we must first define potential terms as strategies. By potential terms, I mean those utterances that are currently meaningless but that could eventually have meaning. Of course, not every utterance will eventually have meaning, and the only ones that will, will be evolutionarily stable strategies. The definition of strategies in terms of language is seemingly the easiest aspect of carrying out this comparison. I will attempt to fill in the blanks; though it will become clear that this is not the most promising route towards charitably interpreting Millikan.

The second step of setting up our evolutionary language game is to define the potential moves that one can make in using a strategy. Since the strategy just is a potential term or utterance, there are only two possible moves: either we use the utterance or we do not. Let us define two strategies, zig and zag.[31] A speaker that adopts the strategy, zig, will either utter "zig" or will not. Similarly, a speaker that adopts the strategy, zag, will either utter "zag" or will not. Zig is an evolutionarily stable strategy if the pay-off for a speaker using zig in a population of other zig users is greater than the pay-off for a speaker using zag in a population of zig users, and *mutatis mutandis* for zag.

Generally, it seems like the pay-off for using a certain strategy could be measured in terms of coordinated speaker use and hearer response. Remember from the discussion of stabilizing functions that we require meaningful devices to have a function that keeps speakers

[31] Here, I am assuming that these are utterances that are currently meaningless.

speaking and hearers responding in standard ways. Perhaps pay-offs can be defined in terms of the stability of the use and response relationships. In this way, we could say that the use of a term is an evolutionarily stable strategy when it has a higher chance of being properly responded to or interpreted.

Despite being able to express a general description of pay-offs for adopting a certain strategy, a major complication arises with further analysis of the moves of the strategies. When two speakers adopt strategy zig, for instance, what this means is that both speakers will utter "zig" or they will not (their decisions are meant to be independent of each other). This does not entail that one zig strategy user will understand the other zig strategy user's utterance of "zig." Remember that, for Millikan, meaning depends on a language device consumer's proper interpretation of an utterance. If a strategy just is an utterance, this says nothing about whether other adopters of this strategy will properly interpret the language device.

Even more strongly, it seems quite likely that the proper consumption of a device is distinct from its proper production. Consider the two users of strategy zig. When user 1 utters "zig," user 2 can either utter "zig" or not. This does not seem to be a system that functions as properly functioning communication systems do.[32] Therefore, it is not clear that we could say that two users adopting the same strategy will have a higher pay-off than those who do not. Since neither adopting the zig nor the zag strategy seems to offer any benefit to a language user, neither of these strategies has the potential to be an evolutionarily stable strategy. In addition, successful language use's dependence upon the understanding of consumers seems to preclude the strategy approach for a game theoretic model of Millikan's theory of language.

Here the reader will surely wonder why we cannot add more moves to the strategy (e.g.,

[32] Admittedly, there are some language devices that do function in this way. For instance, an acceptable response to "hi" would be to say "hi." Still, as this is meant to be a general model for language, this is not an appropriate exchange for the majority of cases.

moves corresponding to interpretation). If we could add correct response moves to a strategy, then it seems like we would be able to conclude that two individuals adopting the same strategy would be said to be more successful than the two individuals who do not. Using this approach, there will be multiple distinct zig and zag strategies. For instance, there can be zig^1, where the user says "zig" or does not and reacts with *a* or does not; and zig^2, where the user says "zig" or does not and reacts with *b* or does not. Similarly, the zag^1 strategy is to say "zag" or not, with or without reaction *a*; and zag^2 strategy is to say "zag" or not, with or without reaction *b*.

Now assume that "zig" has the potential to be interpreted correctly with reaction *a* and "zag" has the potential to be interpreted by an as of yet mentioned reaction, *c*. When there are users of either of the zig strategies in a population, it would be beneficial for language consumers to adopt strategies with reaction *a*. Consequently strategies zig^1 and zag^1 both interpret "zig" correctly; however, there are no strategies that correctly interpret the utterance actions in zag^1. Let us look at this situation abstractly to see how it fits Millikan's view.

Assume that zig^1 is an evolutionarily stable strategy. As a consequence of this, there should be no alternative strategy that can invade. What about zag^1? Presumably, because zag^1 has the correct interpretation moves, it will be used in a proportion of the population. Is zag^1 an evolutionarily stable strategy? First, it is wrong to think of zag^1 as an evolutionarily stable strategy. There is an evolutionarily stable strategy when the device has meaning. Since "zag" is part of the zag^1 strategy, it seems that "zag" would have meaning despite there being an absence of correct interpreters (remember that no strategies have reaction *c*). Still, even though zag^1 is not an evolutionarily stable strategy, it might be possible for it, through its correct interpretative actions, to invade the zig^1 population.

Before answering whether zag^1 could indeed invade the zig^1 population, consider another

case that will aid us in providing the solution. First, consider how strong the evolutionarily stable strategy is and why this is implausible for a language model. As Elliott Sober writes, "a strategy *P* is an ESS [evolutionarily stable strategy] if a population made of 100 percent *P* would be uninvadable."[33] This notion would be a bit extreme when applied to language. If a term had become an evolutionarily stable strategy, then, in a case where an entire population adopted it, no other term could come to replace it. Let us return to the original zig strategy where users either utter "zig" or do not. Imagine that there was a precursor to "zig," "zigocalifragilistic." To say that "zigocalifragilistic" has meaning is to say that it is an evolutionarily stable strategy; and, thus, no other term could come to replace it.

The first question that arises is what it could possibly mean to replace a term that is an evolutionarily stable strategy. To be such a strategy is just to say that the term now has meaning. In order to replace a strategy, we could simply be substituting a new utterance to attach to the old meaning. Perhaps "zigocalifragilistic" means *w*. We introduce the strategy corresponding to "zig;" and, eventually, "zig" becomes an evolutionarily stable strategy that corresponds to the meaning *w*. It does not seem to follow that "zigocalifragilistic" no longer means *w* because a new strategy has successfully "invaded." While new terms arise often enough, the introduction of a new term does not cause its older synonyms to become meaningless. With this in mind, let us return to strategies zig^1 and zag^1.

Remember that I concluded that while strategy zag^1 can correctly interpret the utterance "zig," it was not an evolutionarily stable strategy due to its lacking meaning. Now imagine that there is a strategy zag^3 where the moves are utter "zag" or not and react with c or not. As was previously specified, reaction *c* is the correct interpretation of zag. With the introduction of zag^3, zag^1 can now correctly interpret and be interpreted. If zag^1 becomes an evolutionarily stable

[33] Sober, 139.

strategy, what will happen to zig[1]? Even though their utterance moves are different, perhaps zag[1] could invade zig[1]. If this is the case, then either zig[1] never had meaning or it would become meaningless. Here, one could continue pushing for the strategy method, trying to adapt it and squeeze it to fit Millikan's view. Instead, I propose that there is a much easier interpretation of evolutionarily stable solution that will help us avoid this question altogether.

A more plausible way to think of the meaning of terms is to consider a term to be meaningful when there is an evolutionarily stable *state*. Again, I will approach this possibility by first looking at the notion as it appears in evolutionary game theory and then applying it to the case of language. The path to defining this concept will require more work than defining evolutionarily stable strategies, so I ask the reader to be patient.

I will begin with the components necessary for the definition of an evolutionarily stable state before continuing to the definition itself. Let p and q be states, where a state is a mix of possible strategies.[34] The pay-off of a strategy, i, in state, p, is the fitness; and it is written as $F(i|p)$. We can also think of fitness in terms of a competition between two states where $F(q|p) = \Sigma q_i F(i|p)$: "$F(q|p)$ [is] the average fitness of a group of individuals playing the game against members of a population in state p, when a proportion q_i of the group uses strategy i."[35] Finally, let ε be the proportion of the whole population in state q. With these concepts in place, we can now look at the definition of an evolutionarily stable state: "A state p is called an *ESS* (*evolutionarily stable state*) if for every state $q \neq p$, if we let $\bar{p} = (1- \varepsilon)p + \varepsilon q$ (the perturbed state), then $F(q|\bar{p}) < F(p|\bar{p})$ for sufficiently small $\varepsilon > 0$."[36]

What this definition tells us is that if a population were to adopt the strategies of state q,

---

[34] This can be thought of in the same sense as it was used in Maynard Smith and Price.

[35] Peter D. Taylor and Leo B. Jonker, "Evolutionarily Stable Strategies and Game Dynamics,"*Mathematical Biosciences* 40 (1978), 146.

[36] *Ibid.,* 147.

the fitness of the resulting state (the perturbed state) would be less than the original fitness of state *p*. Note that this definition holds for "sufficiently small" ε, so we will not be considering cases in which a large proportion of the population switches to state *q* from *p*. Using this definition to describe when a term has meaning will help to overcome the problem with the evolutionarily stable strategy approach.

Now, instead of attempting to have a single strategy that encompasses possible term uses and responses, we can have a state that has each of these uses and responses as a separate strategy. Consider the corresponding state for "zig" with strategies s_i where *i* is an integer. Let the even-*i* strategies be speaker strategies and odd-*i* strategies be hearer strategies.[37] Further, if we assume that zig is an evolutionarily stable state, let s_2 and s_1 be the respective speaker uses and hearer responses that correspond to the meaning of the term; and, for the sake of the example, say that they each occur in 2:10 of the population. Now, have a "sufficiently small" proportion of the total population shift to an alternative state, zug. Perhaps in zug, s_2 and s_1 each occur in 1:10 of the population. Since the corresponding state of "zig" is an evolutionarily stable state, it will have a higher pay-off than the perturbed state following the shift. Thus, the state that corresponds to a meaningful term is that preferred state where the possible strategies of use and response are in the most favorable proportions; that is, there is more communicatory benefit to the speakers and hearers in the population in the zig state.

Still, one might find this scenario to be reminiscent of one of the problems that arose for the evolutionarily stable strategies approach. There was an issue with the possibility of a strategy, zag[1], having meaning even though no strategies could correctly interpret its utterance moves. Since the aim of this paper is to distinguish meaningful terms from meaningless

[37] This division is only for the sake of simplicity. I am not committed to there being a one-to-one correspondence between possible speaker strategies and hearer responses.

utterances, the model must surely avoid accidentally labeling the meaningless as meaningful. By stating that there is a state when there is meaning and this state is a mix of all sorts of speaker and hearer strategies, isn't it possible that an utterance that is meaningless will be labeled as meaningful?

This objection can be overcome with a bit of clarification. First, there is an evolutionarily stable state when there is meaning. Second, contra Millikan, the use of the strategy is not an evolutionarily stable solution. When we use a meaningful term, it is the use of a strategy in an evolutionarily stable state (i.e., the evolutionarily stable solution is an evolutionarily stable state). Third, this evolutionarily stable state corresponds to the function category of the language device.

A function category is a way of grouping together things, actions, words, etc. by their proper functions. Millikan defines function categories as being "*essentially* categories of things that need not fulfill their functions in order to have them."[38] Something is a member of a function category in virtue of its ability to be labeled as being defective. This definition may seem counter-intuitive. In order to make it clear, consider an example.

Imagine we're trying to define the members of the function category of a window. Does a wall fit into the same function category as a window? Well, it would seem absurd if, while looking at a wall, someone seriously proclaimed "what a terrible window!" or "your window is defective." Because walls cannot reasonably be considered to be defective windows, walls are not members of the same function category as windows. Thus, we consider something to be a member of a function category when it makes sense to consider it "defective...diseased, malformed, injured, disfunctional, etc.,-hence unable to perform the very functions by which [the

[38] Millikan, "In Defense of Proper Functions," 296.

category] get[s its] name."[39]

In the case of the evolutionary game theoretic model of language, evolutionarily stable states correspond to function categories. From this link, it makes sense to say that when there is meaning, there is an evolutionarily stable state. For a term to belong to a function category, it must have a proper function; and, when there is a proper function, the term has meaning. Now, let us consider how this overcomes the possibility of attributing meaning to meaningless utterances.

Because evolutionarily stable states correspond to function categories defined by the proper function of its members, neither members in the category nor strategies in the state lack meaning. This was considered a problem from the viewpoint of the objector, but I believe that it is a virtue of this model. The strategies that fall into the evolutionarily stable state are just those strategies that can reasonably be said to fail at performing the function for which the category is named. Millikan uses the example of a mispronounced word: "the child that pronounces 'sin' like 'thin' *mispronounces* the word 'sin'...she does not correctly pronounce the word 'thin.'"[40] In my model, this speaker use strategy of uttering "thin" will be in the evolutionarily stable state corresponding to the function category of "sin." The child's utterance of "thin" is meaningful: it means "sin." Still, the strategy is defective. Thus, all of the members of the function category have meaning, but not all of these terms are used or interpreted correctly.

One thing to note here is that this is consistent with Millikan's description of a stabilizing function. As is evident from the preceding example, for a term to have meaning, we need not (and, for Millikan, should not) think of its function as "an invariant function or an average

---

[39] *Ibid.,* 295.
[40] *Ibid.*

function."[41] By accounting for meaning as existing when there is an evolutionarily stable state, we can accommodate the various faulty strategies (i.e., uses and responses) while, also, maintaining a stability in meaning. This is the virtue of the approach.

Finally, from evolutionarily stable states, we can see a way to more clearly differentiate between meaningless utterances and meaningful terms using the history of proper functions. When a term has meaning, there is an evolutionarily stable state that corresponds to the term's function category. Uses and interpretations of the term are strategies in the state. This approach gives us a reason for why this device rather than some other device has the successful history of proliferation that it has. At this point, I have provided the basis of a sketch of what our evolutionarily game theoretic model should look like. An adequate model should at least have meaningful terms correspond to evolutionarily stable states with mixed strategies.

5. Challenges for an Evolutionary Game Theoretic Approach to Language

While following along the previous section, certain concerns may have struck the reader. As I have previously stated, my goal was not to offer a complete game theoretic model of meaning; but, instead, to begin the sketch of how we should view this model. I demonstrated why our evolutionary language game must have terms correlated to evolutionarily stable states rather than evolutionarily stable strategies. In this section, I will point to some of the obstacles that one must overcome in order to quantitatively analyze meaning and to model how a term even comes to have meaning.

It seems that an evolutionarily stable state is most compatible with Millikan's idea of meaningful terms, but this answer requires further elaboration if it is ever to be used in an actual model. Notice that the definition of an evolutionarily stable state relies on fitness. While I offered the general notion of successful communication for defining pay-offs, I have not supplied

[41] Millikan, *Language, Thought, and Other Biological Categories,* 32.

an explicit account of what it means for an utterance or a term to be fit. How could we quantify the fitness of a term in such a way as to be able to use this quantity in evaluations of competitions between states? Obviously the answer to this question requires more work than can fit within the scope of this paper; but, clearly, a notion of fitness is necessary to complete the model.

Once we have a concept of fitness, there is a final question that we must tackle to complete the model: how is the evolutionarily stable state established? Introducing evolutionarily stable states was a first step in an attempt to provide an answer. In order to find a more satisfying response that can help account for when a meaningless utterance becomes a meaningful term, we must turn to dynamics.

An attempt to model the dynamics of a population brings with it its own issues. The main problem is defining pre-quantitatively how language communities interact both within the community and with other communities. If we are not clear on this notion, we will not be able to define what we expect out of the dynamics. For example, Taylor and Jonker base their dynamics off the assumption that each individual always uses the same strategy and offspring use the same strategy as their parents. What changes the proportion of a strategy is that a successful strategy should result in more offspring using that strategy. On the other hand, Martin A. Nowak and Robert M. May use a local interaction model in which strategies within a state only play against the strategies of their immediate neighbors. A winning strategy is adopted by the neighbors; and, in this way, the proportions of strategies within a state change.[42] These are but two of the possibilities for the dynamics of the model. As Taylor and Jonker write, "[t]here are undoubtedly

[42] Martin A. Nowak and Robert M. May, "The Spatial Dilemmas of Evolution," *International Journal of Bifurcation and Chaos,* Vol.3, No.1 (1993): 36.

many ways to do this, each related to certain hypotheses about the population."[43] Thus, without a clarification of population interactions, we cannot plausibly offer the dynamics for how our evolutionarily stable states came to be.

Surely, other issues will arise in formulating an adequate model of meaning acquisition; but, for now, it seems clear that the connection between meaningful terms and evolutionarily stable states has been established. This tells us what will distinguish meaningful devices from meaningless utterances in our model. Future work must be done on clarifying the notion of fitness for language and defining the boundaries and interactions of language communities. Following this, we will be on a much clearer path to establishing an evolutionary game theoretic model of Millikan's teleosemantics.

6. Works Consulted

Lewis, David. *Convention*. Oxford: Blackwell Publishers, 2002.

Maynard Smith, John and George Price. "The Logic of Animal Conflict," *Nature*, 146 (1973): 15–18.

Millikan, Ruth Garrett. "Biosemantics." *The Journal of Philosophy*. Volume 18, Issue 6. (Jun. 1989): 281-297.

---. "In Defense of Proper Functions," *Philosophy of Science*, Volume 56, Number 2 (June 1989): 288-302.

---. *Language: A Biological Model*. Oxford: Clarendon Press, 2005.

---. *Language, Thought, and Other Biological Categories*. Cambridge: MIT Press, 1984.

---. "Speaking Up for Darwin." In *Meaning in Mind: Fodor and His Critics*, edited by Barry M. Loewer and Georges Rey. Cambridge: Blackwell, 1991. 151-164.

[43] Taylor and Jonker, 148.

---. ”Useless Content.” *Teleosemantics*. Eds. Graham Macdonald and David Papineau. New York: Oxford, 2006. 100-114.

Nowak, Martin A. and Robert M. May, “The Spatial Dilemmas of Evolution,” *International Journal of Bifurcation and Chaos,* Vol.3, No.1 (1993): 35-78.

Skyrms, Brian. *Signals: Evolution, Learning, and Information.* New York: Oxford University Press, 2010.

Sober, Elliott. *Philosophy of Biology*. 2nd ed. Boulder: Westview, 2000.

Taylor, Peter D., and Leo B. Jonker. "Evolutionarily Stable Strategies and Game Dynamics." *Mathematical Biosciences* 40 (1978): 145-56.