

BIOSEMANTICS AND MAGNETISM AS METAMETASEMANTIC THEORIES

Filip Kawczyński¹ 

¹ University of Warsaw, Poland

Review paper – Received: 31/10/2024 Accepted: 14/09/2025

ABSTRACT

Semantics assigns meanings to words and phrases, providing a framework for understanding language. Metasemantics, on the other hand, seeks to explain the origin of these meanings, exploring why certain meanings are ascribed to particular words, what does it even mean “to mean something”, or what is the ontological status of meanings. Although semantics and metasemantics may appear to provide a complete picture of language, it has been suggested that another theoretical level—above metasemantics—is necessary for our theories to avoid (mainly) the problem of indeterminacy. This additional level is called metametasemantics. This article examines two leading candidates for metametasemantics: Lewis’s Reference Magnetism and Millikan’s Biosemantics. While both theories successfully anchor metasemantics in non-linguistic facts to address the problem of indeterminacy, my analysis reveals crucial differences in their explanatory power and theoretical robustness. The paper develops a four-criteria framework for evaluating metametasegmentic theories, distinguishing between threshold conditions (indeterminacy resolution) and gradable criteria (empirical grounding, theory comparison, and pluralism accommodation). Through systematic comparison, I demonstrate that while Magnetism provides an elegant solution to indeterminacy through natural properties, its reliance on the notion of “naturalness” creates significant theoretical challenges. Biosemantics, grounded in evolutionary biology, offers more rigorous criteria for adjudicating between alternative metasegmentic theories while maintaining flexibility across linguistic domains. The analysis pays particular attention to how each theory handles persistent issues like functional indeterminacy and theoretical underdetermination.

Keywords: biosemantics; semantics; metametasemantics; indeterminacy.

1. Introduction

The study of meaning in language, traditionally the domain of semantics, has evolved significantly over time. Semantics assigns meanings to words and phrases, providing a framework for understanding language. These days we are witnessing the process of “redistribution” of the problems that once were considered to belong to the philosophy of language—a significant number of those problems have been taken over by disciplines like psychology, cognitive science, or linguistics. And since there are still, without a doubt, many philosophical concerns about language, the term “metasemantics” has been gaining popularity.

Metasemantics seeks to explain the origin of these meanings, exploring why certain meanings are ascribed to particular words, what it even means “to mean something”, or what is the ontological status of meanings (Burgess and Sherman 2014). In this sense, general metasemantics covers many of the traditional problems in the philosophy of language.

On the methodological side, metasemantics is crucial as a justification of semantics. Let’s take, for instance, a semantic theory of inferentialism, which can “calculate” the meaning of any phrase by showing its place in a network of interrelated inferences. We can easily imagine a scenario in which it turns out that a competing semantic theory—for instance verificationism or even some more eccentric theory based on astrology—“calculates” meanings in a way that yields the same results as inferentialism. Both theories fulfill their semantic task equally effectively. Metasemantics is supposed to offer a basis for choosing between alternative semantic theories.

To illustrate the distinction between semantics and metasemantics, we will use a well-known example involving proper names and definite descriptions. When considering the relationship between proper names and definite descriptions, two major positions are typically distinguished. First, one may adopt descriptivism, according to which a proper name is semantically equivalent to a definite description—the semantic value of the description determines the semantic value of the name. This is a *semantic* thesis: it aims to specify what the semantic value of a proper name is. On the other hand, one might hold that a description does not constitute the semantic value of the name, but merely fixes its reference. If one takes the reference to be the semantic value of the name—as is standard in theories of singular propositions—this yields a *metasemantic* position: it aims to explain where the semantic values of proper names come from. If a proponent of such a version of the singular propositions framework sought to incorporate Saul Kripke’s (1980) notions of initial baptism and

causal chains, then any explanation that appeals to these Kripkean notions would likewise be *metasemantic* in character—for they do not tell us what the semantic value of a name is, but rather what it is grounded in.¹

It might thus be said that semantics is supposed to provide a general picture of how language works, while metasemantics is expected to answer the question of why we can assume it works in this way rather than in another.

However, probably the biggest challenge that every metasemantic theory has to face is the notorious problem of indeterminacy. Indeterminacy within metasemantics (as well as in semantics) consists in ambiguity and lack of determinacy in how meanings are assigned and justified. This problem arises due to the complex and multifaceted nature of language, where multiple interpretations can satisfy the same set of linguistic data. Indeterminacy challenges the very foundations of metasemantic theories, calling into question their ability to provide definitive explanations for meaning.

Introducing a metametasemantic level is supposed to solve the problem of indeterminacy affecting metasemantic theories. However, there is still no agreed-upon view on what requirements a good metametasemantic theory should satisfy, or on how to adjudicate between competing theories. The goal of this paper is to offer a set of requirements and to use them to assess prominent metametasemantic theories.

The paper is structured as follows. In Section 2, I discuss the key problem of metasemantics i.e. indeterminacy, and explain how the metametasemantic level helps address this issue. I also formulate a list of requirements that a good metametasemantic theory should meet. In Section 3, I test Lewis's Magnetism as the first candidate for a metametasemantic theory. In Section 4, I subject Millikan's Biosemantics to the same test and argue that, although designed with a different intention, Biosemantics performs well as a metametasemantic theory—particularly in resolving the key problem of indeterminacy. In the final section, I summarize the earlier findings by comparing Magnetism and Biosemantics in their role as metametasemantic theories and present the final conclusions.

¹Another example is Robert Brandom's (1994, 2000) semantic theory of inferentialism, which is complemented by a metasemantic foundation provided by his theory of reasons and social practices.

2. Indeterminacy and Metametaseantics

Theodore Sider (2011), based on Hilary Putnam's (1981) model-theoretic argument, indicates that each metasemantic theory faces the problem of indeterminacy. Putnam (1981) argues that model-theoretic considerations show that any theory expressed in predicate logic consistent with all available constraints admits multiple models with systematically different reference assignments. This leads to the issue of the indeterminacy of reference, or more generally, of interpretation for a language.

Putnam's argument highlights a general issue. Any metasemantic theory, in order to perform its key function of grounding the semantics of a language, needs to solve the problem of indeterminacy and to show which of the many (possibly infinitely many) interpretations or meaning assignments for that language is the correct one. Indeterminacy will manifest differently depending on how a given metasemantic view is constructed. For example, in the case of Brandom's inferentialism, the question arises as to how practices are to be individuated and which elements of those practices are crucial for meaning determination. Likewise, if one adopts Ruth Millikan's (1984, 1989, 2000) biosemantic theory, which specifies general conditions that a relation must satisfy in order to count as a reference relation, the question remains as to which of the many relations satisfying those conditions are the correct ones.

Given this issue, Sider postulates that above metasemantics, a further level—namely metametaseantics—is necessary to ultimately dispel the specter of indeterminacy. The metametasemantic level is supposed to allow us to discriminate between alternative metasemantic theories and the interpretations of language they imply. Metametaseantics is thus intended not so much as a theory explaining the functioning of language, but rather as an anchor that prevents indeterminacy.²

The question, then, is how to understand this anchoring relation. It cannot plausibly consist in a set of semantic laws, since such laws would themselves be subject to indeterminacy.³ To eliminate this indeterminacy, it seems necessary to anchor a metametasemantic theory in a domain other than the linguistic one—namely, a domain that is not itself burdened by indeterminacy. This, in turn, leads us to search for the appropriate non-

² The concept of metametaseantics faces methodological resistance. For instance, Simchen (2017) argues that accepting metametaseantics undermines the explanatory power of metasemantic theories, thereby rendering it irrelevant which metasemantic theory we adopt—whereas it evidently does matter. Kawczyński (2021), by contrast, argues that Simchen's argument is invalid and offers further reasons for dispelling the suspicion that metametaseantics renders metasemantics irrelevant.

³ See Miller (2008) for the debate on Chomsky's argument against Quine's indeterminacy thesis.

linguistic states of affairs capable of grounding metametase-mantic theories.

An important condition imposed by this perspective is that a metametase-mantic theory must be compatible with non-linguistic empirical data and with the theories describing those data. For example, if a metametase-mantic theory anchors language in mental states, it must be consistent with empirical data on mental states, as well as with psychological, neurobiological, and cognitive theories of those states. Compatibility with such theories thus provides one of the criteria that metametase-mantics can use to assess competing metase-mantic theories or interpretations of language.

a. Metametase-mantics: The Checklist

With this in mind, we can now identify the general expectations that a metametase-mantic theory should meet. Three such expectations have already emerged. First and foremost, metametase-mantics is expected to shield metase-mantics from indeterminacy. Second, it is supposed to do so by appealing to non-linguistic facts. Third, metametase-mantics should be capable of comparing competing metase-mantic theories. This third requirement gives rise to a further one: metametase-mantics must provide a sufficiently open framework for metase-mantic theories and remain continuous with them. Without such continuity, metametase-mantics would be unable to compare metase-mantic theories at all. In other words, metametase-mantics ought to “speak the same language” as metase-mantic theories—that is, the core theoretical terms or concepts employed by metase-mantic theories must be interpretable within a metametase-mantic account.

Another point, which is not a necessary condition but is, in my view, a highly desirable feature, is that metametase-mantics should allow for the application of different metase-mantic theories to different linguistic phenomena. A common issue in the philosophy of language is that, at the metase-mantic level, there is a default tendency to construct a single, monolithic theory for an entire language. This challenge is particularly difficult because language is an extraordinarily internally diverse phenomenon: some of its parts are highly context-sensitive, while others are not; in some domains truth is the central category, whereas in others it plays little or no role, and so on. It is well known, for example, that some theories are well suited to constatives but fail to account adequately for performatives, or that concepts which successfully explain aspects of everyday communication cannot straightforwardly serve as accounts of the

literal meanings found in dictionaries. Examples of such mismatches could be multiplied indefinitely.

It seems that one of the primary motivations for philosophers in seeking a monolithic metasemantic theory is the desire to avoid indeterminacy. A situation in which different metasemantic theories are applied to different areas of language potentially generates more possible interpretations than the application of a single overarching theory, partly because the boundaries between the domains of different theories may be blurred. However, if metametasemantics—by appeal to non-linguistic facts—can effectively mitigate the threat of indeterminacy, then the prospect of applying different metasemantic theories to different linguistic phenomena appears far more feasible.

Let us now summarize what has been established regarding the conditions for a metametasemantic theory. An adequate metametasemantic theory is one for which the answer to the following questions is “yes”:

- **MMS1.** Does it resolve indeterminacy?
- **MMS2.** Does it anchor metasemantics in non-linguistic facts (and, furthermore, is it compatible with scientific theories concerning those facts)?
- **MMS3.** Does it provide tools for comparing metasemantic theories (and, more generally, a framework within which metasemantic theories can be compared and with which they are continuous)?
- **MMS4.** Does it allow for the application of different metasemantic theories to different linguistic phenomena?

While all four criteria carry philosophical significance, they do not play an equal role in evaluation. I maintain that MMS1 (resolving indeterminacy) functions as a binary necessary condition—any theory that fails to meet this threshold requirement cannot be seriously considered a viable metametasemantic candidate. Crucially, MMS1 functions as a discrete, all-or-nothing criterion rather than admitting degrees of satisfaction: either a theory successfully eliminates indeterminacy by uniquely determining correct interpretations in all cases, or it fails this test entirely. There exists no intermediate position where a theory might partially fulfill MMS1.

The remaining criteria, by contrast, appear to function as scalar measures. They admit gradable evaluation, allowing theories to satisfy them to varying degrees of theoretical adequacy. Where MMS1 serves as a gatekeeping condition, MMS2-MMS4 permit comparative assessments of how satisfactorily different theories address their respective demands.

In the subsequent sections, I will compare two candidate metametasemantic theories in light of the criteria outlined above. The first candidate is David Lewis's Magnetism, and the second is Ruth Millikan's Biosemantics. I argue that both theories fulfil the fundamental task of metametasemantics, namely avoiding indeterminacy. However, I will endeavor to demonstrate that there is at least one respect in which Biosemantics performs slightly better than Magnetism as a metametasemantic theory.

3. Candidate 1: Reference Magnetism

Lewis's theory of Reference Magnetism posits that the terms we use in language naturally align with certain properties in the world because some properties are more "natural" than others. These natural properties act as "magnets" that attract the reference of our terms, leading to an independent way of connecting language to the world. According to Magnetism, the correct interpretation of language, among many possible ones, is the one that not only maximizes the truth of our statements but also aligns the references of words with the most natural properties, or, as Lewis puts it, "respects the objective joints in nature" (Lewis 1984, 227). Specifically, when evaluating two competing interpretations of a particular predicate, the interpretation that more accurately refers to the objective natural properties in the world is deemed to be the more appropriate one.⁴

Lewis proposed this theory precisely to avoid indeterminacy. The theory grounds reference in objective, mind-independent reality by linking it to natural properties, offering a coherent framework for understanding how words latch onto the world. In essence, Lewis's Magnetism seeks to provide a systematic way to think about reference and meaning in terms of metaphysical realism, ensuring that our language is more closely tied to the objective structure of reality. It is difficult to deny that Magnetism successfully addresses the problem of indeterminacy: regardless of how many and what kinds of language interpretations are proposed, it allows us to determine one as the correct one, guided by the criterion of naturalness.⁵

⁴ For example, the property of being green for an entire existence is more natural than the property of being green until a certain point in the future and then turning blue. Therefore, the former property is a more suitable candidate for defining the extension of the predicate "green". Consequently, it allows for choosing the standard interpretation of "green" over the "grue/bleen" interpretation.

⁵ Of course, at a given moment, we may not be able to determine which interpretation is more natural, but the crucial point is that one of them undoubtedly is more natural. In the past, the predicate "golden" was used to describe both gold and pyrite. Eventually, it became clear that these are two different substances, leading to the addition of the predicate "pyritic". A proponent of Magnetism can explain why a language containing only the predicate "golden" has a defective semantics: namely, because such a language carves nature at its joints less successfully than a language that includes both "golden" and "pyritic". It is clear, however, that a magnetist explanation can arise only once people have actually discovered the differences between gold and pyrite.

a. Minor Problems for Magnetism

Magnetism is quite an innovative philosophical approach, as is often the case with ideas developed by Lewis. Unsurprisingly, it has also been met with significant criticism. There are several common objections to Magnetism:

1. Magnetism oversimplifies the complexity of real-world reference by ignoring various contextual factors, social practices, and pragmatic considerations. The theory fails to capture the nuances of everyday language, such as speaker intentions and listener interpretations (see, e.g., Schwarz 2014).
2. Magnetism does not adequately explain polysemy and homonymy. It has been pointed out that words like “bank” can have multiple related or completely different meanings, which the theory’s focus on natural properties does not address (see, e.g., Field 1980; Schiffer 2003). This indicates a lack of flexibility needed to handle diverse linguistic phenomena.
3. Magnetism aligns neither with externalism nor internalism—it straddles both approaches without fully committing to either, undermining its theoretical coherence (see, e.g., Wilson 1982; Wright 1992).
4. Magnetism seems to ignore the dynamic aspect of language. Language evolves over time, influenced by cultural shifts, technological advancements, and social norms. Magnetism’s emphasis on fixed natural properties does not accommodate this fluid nature, leaving it silent on how linguistic change should be understood and integrated into its framework (see Schwarz 2014).

b. Major Problems with Magnetism

The above objections are valid and significant, but in my view, they do not constitute the most substantial problem with Magnetism. The aim of this discussion is not to defend Magnetism, and I will not delve into the details of how this theory can be defended against the aforementioned objections.⁶ It seems that the aforementioned problems of Magnetism pale in comparison to the benefits it provides—namely, the resolution of the indeterminacy problem. In other words, accepting these issues appears to

⁶ For that see Kovacs (2020), Manley and Wasserman (2008), Schaffer (2009), Sider (2011), Weatherston (2013), Williams (2007), among others.

be a price worth paying for dealing with indeterminacy. The strongest objection to Magnetism is that it is often arbitrary, *ad hoc*, and overly speculative, with a vague and almost “magical” character, especially when it comes to defining its core concept, namely *naturalness*.

Lewis (1983) introduces “naturalness” as a primitive concept. The concept, however, is not considered by Lewis as lacking any content, but as a matter of fact as a theoretically robust concept. While resisting reductive definition, naturalness plays three fundamental roles. First, natural properties serve as the objective basis for genuine similarity relations between objects. For instance, all gold specimens share identical microstructure (atomic number 79), whereas pyrite (FeS_2) exhibits a distinct atomic configuration. This microphysical difference makes “being gold” a more natural property than the disjunctive “being gold-or-pyrite”.

The second role of natural properties is that they are both (a) causally efficacious and (b) appear in the fundamental laws of completed science. This dual requirement ensures that naturalness tracks properties with genuine explanatory power in scientific practice, not merely conventional classifications.

Thirdly, Lewis metaphorically describes natural properties as “carving nature at its joints”. This entails that they reflect mind-independent divisions in reality’s structure, which scientific investigation progressively reveals. It should be emphasized that, according to Lewis, science—particularly the “final physics”—*reveals*, and does not stipulate, constitute, or postulate, those properties. The naturalness of e.g. “gold” derives from objective atomic structure, not semantic or theoretic conventions. Even without absolute metrics, it is a largely uncontroversial assumption—based on comparative ‘phenomenological’ judgments—that monadic properties (e.g., “gold”) are more natural than disjunctive ones (e.g., “gold-or-pyrite”).

This framework is supposed to provide an objective basis for reference determination while respecting science’s epistemic authority. The theory’s alleged power lies in its ability to explain referential stability without appealing to mysterious semantic relations—natural properties ground reference through their role in the causal-nomological structure of reality itself.

However, several scholars have found Lewis’s explanation of naturalness unsatisfactory. For example, Hartry Field (1980) and Mark Wilson (1982) argue that without a clear and rigorous definition of what makes a property “natural”, the theory becomes speculative. It relies on an intuitive sense of

naturalness that varies significantly among different contexts and individuals.

Scott Soames (2005) contends further that the selection of natural properties appears *ad hoc*, as it often depends on subjective judgments rather than objective criteria. This arbitrariness undermines the theory's claim to provide a non-arbitrary grounding for reference. Soames argues that in the absence of objective standards for naturalness, the theory risks circular reasoning, where natural properties are defined by their ability to attract reference, which in turn is explained by their naturalness. Furthermore, Crispin Wright (1992) and Huw Price (2011) highlight the *ad hoc* nature of the theory as they argue that Lewis's appeal to natural properties often seems like a convenient way to bypass more rigorous and detailed semantic analysis. By invoking natural properties, the theory sidesteps the complexities involved in explaining how language functions in diverse and dynamic contexts, making it appear more as a speculative shortcut than a substantive explanation. Stephen Schiffer (2003), on the other hand, points out that the vagueness of naturalness leads to practical difficulties in applying the theory. Since the concept lacks precise boundaries, it is challenging to determine which properties qualify as natural in specific instances.

More recently, Cristina Nench (2017) exposed a tension between Lewis's use of natural properties and his anti-essentialist counterpart theory. Although Lewis employs naturalness to ground similarity relations between counterparts, Nench shows that this move inadvertently smuggles in essentialist commitments. Even if certain properties (e.g., mass or charge) are deemed "natural", their naturalness does not entail metaphysical necessity—a property's status as objectively similarity-determining does not make it essential to an object's identity across possible worlds. Furthermore, the selection of which natural properties to privilege remains arbitrary, as no non-question-begging criteria exist to favor, say, mass over charge as the basis of counterpart relations. This arbitrariness undermines Lewis's attempt to reconcile natural properties with his anti-essentialist framework, revealing a deeper inconsistency in his metaphysics of modality.

A related objection is raised by Michela Massimi (2017) in her discussion of Lewis's Best System Account (BSA) of laws. She argues that naturalness is not only theoretically unstable but also empirically inadequate for capturing scientific practice. Lewis's BSA assumes that laws emerge from a balance of simplicity and strength in a system grounded in natural properties, but Massimi shows that judgments of naturalness are perspective-dependent: what counts as a "natural" property

in classical mechanics (e.g., position) may not hold the same status in quantum field theory (e.g., spin). This variability reflects a broader epistemic relativity: different scientific communities legitimately prioritize different theoretical virtues, leading to competing but equally valid “natural” descriptions of phenomena (e.g., thermodynamic vs. statistical mechanical accounts of temperature).⁷

These critiques reveal a systematic weakness in Lewis’s reliance on naturalness: whether applied to reference, modality, or laws, the concept fails to provide the objective, theory-neutral grounding he claims, instead collapsing into circularity, arbitrariness, or perspectival relativity. While Lewis’s framework remains influential, these challenges suggest that any viable successor must either radically reconceive naturalness or abandon it in favor of a more epistemically and metaphysically nuanced alternative.⁸

c. Magnetism as metametasemantics

Let us now put Magnetism through our checklist of conditions which an adequate metametasemantics should meet.

- **MMS1.** Does it resolve indeterminacy?
 - **YES.** With ease.
- **MMS2.** Does it anchor metasemantics in non-linguistic facts (and is compatible with those facts as well as scientific theories concerning those facts)?
 - It is a “YES, but...” answer in this case. Undoubtedly, Magnetism anchors any metasemantics in non-linguistic facts concerning natural properties. It is compatible with any data about the world and any scientific theory of the world. However, therein lies its weakness. As a theory that can

⁷ Massimi proposes replacing Lewis’s monolithic BSA with a Robustly Best System model, which accommodates pluralism by recognizing multiple contextually optimal systems, each reflecting distinct but functionally salient explanatory goals. This shift from absolute naturalness to contextually embedded naturalness better aligns with the actual history and practice of science, where explanatory frameworks evolve and compete without converging on a single privileged description of reality.

⁸ Some philosophers have attempted to refine Lewis’s notion of naturalness to address concerns about vagueness and arbitrariness. For instance, J. R. G. Williams (2007) suggests a more epistemically modest account of natural properties, emphasizing eligibility as a pragmatic tool rather than a metaphysically privileged category. Similarly, James Ladyman and Don Ross (2007) in their work on “ontic structural realism” argue that the fundamental structure of reality may itself undercut the traditional notion of properties as basic natural units, suggesting that naturalness must be reconsidered in more relational or structural terms. While these efforts enrich the debate, they also highlight the underlying instability of the concept of naturalness in Lewis’s original framework.

“accommodate everything” it loses much of its explanatory power. It can be said that Magnetism fulfills the condition in question, but it is a “vacuous fulfillment”.⁹ One could argue that until we reach the stage of “the final physics”, all resolutions concerning naturalness may prove to be mistaken in the sense that they are dictated by the adopted conceptual framework, and thus by a particular interpretation of language.¹⁰

- **MMS3.** Does it provide tools for comparing metasemantic theories (and a framework for metasemantics that is continuous with them)?

- This time, the answer is “Technically YES, but practically NO”. Magnetism will undoubtedly favor naturalistic theories over non-naturalistic ones and, in this sense, has the capacity to compare theories. Certainly, Magnetism is also continuous with all realistic empirical theories. However, when the scope of metasemantics is restricted to naturalistic theories, Magnetism lacks the means to indicate which of two theories (yielding the same semantic results) is better. And the reason for this lies in the flaws of Magnetism summarized above under The Major Problem. Magnetism’s response to the question of which interpretations/theories to choose is very general and, as have been pointed out, unclear. If Magnetism demanded more from reference than just the vaguely defined naturalness, it might be possible to point out certain advantages, say, of descriptive interpretationism over the theory of causal communicative chains. However, since Magnetism does not do that, any metasemantic theory, as long as it is sufficiently naturalistic, can fit within its framework equally well and this deprives us of a tool for choosing between alternative theories.

- **MMS4.** Does it allow for the application of different metasemantic theories to different linguistic phenomena?

⁹ The reason why I answered “Yes” in MMS1, but “Yes, but...” in MMS2 is the following: Magnetism has an internal problem with the notion of naturalness, whereas it has no issue at all with solving the indeterminacy problem. If Magnetism was updated and offered a satisfactory characterization of naturalness, it would effectively eliminate indeterminacy. On the other hand, Magnetism faces the difficulties in fulfilling the task formulated in MMS2. No possible revision of the theory of Magnetism itself can make it better suited to perform the function assigned to it in MMS2.

¹⁰ Obviously, what is at issue here are Quine’s classic skeptical arguments concerning the impossibility of an objective, language-independent evaluation of scientific theories, to which the equally classic response—one that can be used to defend Lewis—is the argument from the success of science.

- **YES.** As a very permissive theory, Magnetism does not impose any restrictions on applying different theories to different linguistic phenomena.

Overall, Magnetism performs quite well on the proposed criteria—primarily, because it addresses the problem of indeterminacy by relying on non-linguistic factors, and it provides significant flexibility for metasemantic solutions. At the same time, it may be said that this flexibility stems not so much from the versatility of Magnetism, but from the lack of precise definition of the notion of “naturalness”.

4. Candidate 2: Biosemantics

The second candidate for playing the role of metametasemantics is Millikan’s (e.g., 1984, 1989, 2000) Biosemantics. According to this account, the meanings of mental states and linguistic items are rooted in their biological functions, which are determined by evolutionary processes. These functions provide the basis for reference and representation, allowing for the possibility of both accurate representation and misrepresentation.

The key notion of the theory is the notion of proper function. Proper function of a system is whatever ancestral systems did which caused systems of that type to be preserved and/or proliferated in the population. Meanings are, at bottom, messages from producers to consumers, and meaning is determined not by the function of producers but by the functions of consumers—the systems that use the meaning to perform their proper functions.

Biosemantics provides an unprecedented explanation of linguistic phenomena in philosophy using scientific concepts—specifically, those from biology and evolutionary theory. Thus, similar to Magnetism, Biosemantics offers a very innovative approach to language. And, likewise, this theory has faced numerous objections.

a. Standard Objections to Biosemantics

Critics argue that Biosemantics’ heavy reliance on evolutionary theory raises questions about its applicability to novel situations that were not present during the trait’s evolution. In other words, historical basis of proper functions may not provide clear guidance in new and unprecedented contexts (see, e.g., Fodor 1990; Neander 1991). Additionally, some traits

may be byproducts of other adaptations (spandrels) rather than having a direct evolutionary function, complicating the determination of proper functions and their role in meaning (see Gould and Lewontin 1979). Moreover, traits can change their functions over time. In this regard, an individual's mental capacities may have evolved for one purpose but are now used for very different functions in contemporary settings, raising questions about the stability and applicability of proper functions as determinants of meaning.

These problems spill over into issues for Biosemantics. If the functions of mental states can evolve and change, it becomes difficult to anchor meanings in their original evolutionary purposes (see Cummins 2002). Moreover, it has been argued that Biosemantics struggles to explain the meaning or reference of states in artificial systems, such as computers, which are not products of natural selection. This limitation suggests that Biosemantics may not be universally applicable (see, e.g., Stich 1992).

Apart from Millikan herself, Biosemantics has been defended against these objections by several of its supporters (see, e.g., Dretske 1988; Neander 1991, 2002; Papineau 1987). Similar to the case with Magnetism, I do not delve into the details of these arguments in this paper. However, I would like to present one consideration that addresses these objections at a methodological level and, I believe, strengthens the position of Biosemantics as a candidate for being a metametasemantic theory.

In my opinion, the doubts related to anchoring Biosemantics in evolutionary theory can be roughly summarized by the question: What would happen to Biosemantics if evolutionary theory turned out to be false? Certainly, like any other scientific theory, evolutionary theory is neither irrefutable nor complete, and therefore it is undoubtedly expected to undergo further minor or major reformulations. However, changes in evolutionary theory regarding specific organs, species, or entire ecosystems will not affect Biosemantics, because what Biosemantics needs from evolutionary theory are only very general concepts, such as adaptation and survival, as well as the notion of proper function constructed on their basis. It seems that Biosemantics would face greater difficulties if the entire theory of evolution were rejected, along with the belief in the existence of mechanisms such as natural selection and adaptation. This scenario is highly unlikely, given that evolution by natural selection is a long-standing theory with a substantial body of data supporting it. Nonetheless, logically speaking, such a scenario is, of course, possible. In such a situation, Biosemantics would be on par with Magnetism. There is no empirical theory of naturalness—and just as Lewis postulates naturalness without having any theory of naturalness, so too

could Millikan, without the theory of evolution, postulate the occurrence of adaptive processes that anchor metasemantics. In such a situation, both theories would fulfill the tasks of metametasemantics but would also be burdened by the charge of arbitrariness. Even if this extremely unlikely scenario occurred in which the basic assumptions of the theory of evolution were rejected, Biosemantics would perform at least as well as Magnetism in the field of metametasemantics.

b. The Major Objection to Biosemantics

Perhaps the most significant issue facing Biosemantics is the notorious problem of indeterminacy. This issue is of utmost interest to us in the context of seeking a robust metametasemantic theory. In this case, indeterminacy takes the form of *functional* indeterminacy—as presented by Neander:

Motivated toads will hunt and try to eat anything with a suitably worm-like configuration: i.e., they will hunt and try to eat anything that's, roughly, small, elongated and moving parallel to its longest axis. They cannot discriminate between toad food and cardboard cutout rectangles with the right configuration, but in their ancestral environment, things with that configuration were often enough toad food, so a device that responded to these features sufficed for the job. We can describe the function of the relevant part of the toad's perceptual system in different ways: e.g., as detecting toad food and as detecting things with the right configuration of features. What does the toad's perceptual representation represent then? Does it misrepresent the cardboard cutout as a toad food? Or does it correctly represent the cardboard cutout as an item in worm-like motion? (Neander 2006, 386)

Fodor (1990) addresses the same problem from a slightly different angle as he points out that natural selection cannot discriminate between co-extensive features. We can think of the following example: if it is adaptive to run when you hear someone saying (A) “a creature with a heart”, it is also adaptive to run when you hear someone saying (B) “a creature with kidneys”. Thus, we can describe your proper function equally well as “run when hearing (A)” and “run when hearing (B)” —so the function is not determinate.

In short, we deal here with the classic indeterminacy problem—ascriptions of proper functions are too indeterminate themselves to be able to ascribe meanings in a determinate way. Millikan, Papineau or Neander have

developed defenses of Biosemantics against arguments of indeterminacy. The discussion between them and the critics of Biosemantics has not led to a definitive conclusion, and the specter of indeterminacy still looms over Biosemantics. In the following sections of this discussion, I aim to defend Biosemantics as a strong candidate for metametasemantics. To give my proposal the greatest possible strength, I will assume the unfavorable thesis for Biosemantics that it is indeed burdened with the functional indeterminacy characterized as above.

c. Biosemantics as Metametasemantics

Assuming that Biosemantics suffers from the functional indeterminacy, we must acknowledge that it is thereby disqualified as a ground for a good semantic or metasemantic theory. My main thesis is that, despite this, Biosemantics shows great promise as a metametasemantic theory. To justify this, I will subject Biosemantics to the test of the four requirements for metametasemantics. The most interesting question, naturally, is whether Biosemantics, despite the functional indeterminacy, can provide metasemantics with protection against indeterminacy. This issue requires a broader commentary; therefore, I will dwell on it for a while before moving on to the remaining conditions of the checklist.

i. Biosemantics vs Indeterminacy

In my view, the answer to the question of whether Biosemantics protects metasemantics from indeterminacy is affirmative, and the simplest explanation is as follows: Biosemantics anchors metasemantics not in the *theory* of evolution but in the process of evolution itself—in evolutionary *facts*. The biosemantic explanation relies on the concept of proper function, which is a concept derived from the evolutionary theory; identifying proper functions is indeed fraught with indeterminacy, and as such, assigning meanings to words (i.e. semantics) or explaining the origins of meanings (i.e. metasemantics) is also indeterminate. However, these are not the tasks of metametasemantics, which aims to anchor metasemantics in non-linguistic facts, as such facts are inherently not indeterminate. And Biosemantics fulfills this task by anchoring metasemantics not in the indeterminate theory of evolution, but in the determined evolutionary facts.

Let us examine once again how Magnetism—considered as a theory that excels in this matter—dispels the threat of indeterminacy. We can return to the well-known scenario where we are dealing with two interpretations of a language, each backed by a reasonable metasemantic explanation:

- Interpretation (I1) includes only the predicate “golden”;

- Interpretation (I2) includes two predicates: “golden” and “pyritic”.

Magnetism chooses (I2) over (I1) because (I2) better fulfills the requirement of naturalness. In other words, it is more correct to have two predicates, since they better *carve nature at its joints* than just one predicate. The reason for that hinges on the interplay between microstructural facts and comparative naturalness judgments. The atomic structure of gold (79 protons) objectively unifies all gold specimens as a natural kind, whereas pyrite (FeS_2) exhibits a distinct crystalline configuration. This microphysical divergence renders “being gold” a metaphysically privileged property compared to the disjunctive “being gold-or-pyrite”—not by linguistic convention, but because gold’s structural uniformity features more directly in fundamental laws and grounds genuine similarity relations. Crucially, Lewis’s framework permits such comparative assessments even without absolute naturalness metrics: monadic properties consistently outperform disjunctive ones in law-like centrality and causal robustness, irrespective of human categorization practices. Thus, when reference appears indeterminate (e.g., whether “gold” denotes Au or AuVFeS_2), the theory appeals not to interpretive preferences but to nature’s own graded structure—where gold’s microphysical unity objectively outweighs the gerrymandered disjunction.

Biosemantics operates similarly in this realm. Among alternative interpretations, one is the most conducive to survival, and it is precisely this interpretation that is the correct one. If we cannot identify the proper functions responsible for making a particular interpretation better in this respect—whether due to incomplete data or the indeterminacy of evolutionary theory—it remains a fact that a given interpretation, roughly speaking, is evolutionary more efficient than others. Evolution itself—just like naturalness, if we believe in it—is not indeterminate; it is a set of facts that occur independently of the shortcomings of our theory aspiring to explain these facts.

If Biosemantics were to play the role of metasemantics, the indeterminacy of proper functions would be a serious problem because there would be several alternative explanations, involving different proper functions, for a single phenomenon of a given word having a certain meaning. Biosemantics engaged in the role of metametasemantics, however, is not obliged to indicate which specific proper function is responsible for a particular meaning. For Biosemantics to fulfill its task as metametasemantics it is enough to claim that *some* proper function is responsible for the meaning.

The biosemantic account of meaning determination is uniquely resistant to challenges of functional indeterminacy precisely because it is grounded in

the dynamic, historically contingent processes of biological adaptation. Unlike theories that rely on static metaphysical posits, Biosemantics recognizes that evolutionary mechanisms inherently resolve apparent symmetries between competing functions through concrete biological pressures. When confronted with hypothetical cases of perfectly balanced adaptive functions—where two traits appear equally fitness-enhancing—the biosemantic framework appeals to the following realities of evolutionary systems.

First, biological selection rarely operates in idealized equilibrium. Even when two functions seem equally advantageous in a given environment, subtle second-order factors—developmental constraints, pleiotropic effects, or shifting ecological conditions—inevitably break the symmetry over time. What might appear as referential indeterminacy at a snapshot moment is resolved through the cumulative, historical trajectory of selection pressures. The biosemantic criterion for meaning is not an instantaneous functional equivalence but the long-term stabilization of interpretative norms through sustained adaptive success.¹¹

Second, Biosemantics does not require perfect determinacy at every moment, only that evolutionary processes tend toward stable functional outcomes. Temporary ambiguities in function assignment (such as transitional periods in trait evolution) do not undermine the theory, because meaning is anchored in the broader statistical tendency of selection to favor interpretative consistency where it enhances reproductive success.

Finally, the resilience of Biosemantics does not hinge on the irrevocable truth of any particular evolutionary model (e.g., neo-Darwinism). Rather, it depends on the more general principle that functional organization in living systems—however ultimately explained—must necessarily generate stable semantic norms. Even if future science were to radically revise our understanding of heredity or adaptation, any plausible alternative would still need to account for how functional behaviors become entrenched in populations. Biosemantics could seamlessly integrate such advances, as its core insight concerns the grounding of meaning in objective biological functions, not the specific machinery of natural selection.

By situating meaning within the empirically tractable, self-correcting processes of evolution, Biosemantics avoids the static indeterminacy that plagues theories dependent on *a priori* metaphysical distinctions. Where competing accounts struggle to explain why reference resolves one way rather than another, Biosemantics points to measurable, historically

¹¹ I am grateful to Reviewer 2 for their valuable insight on this significant point.

emergent facts about functional success—making it uniquely equipped to handle both hypothetical challenges and real-world complexities of meaning determination.

Thus, in terms of resolving indeterminacy, Biosemantics performs no worse than Magnetism,¹² and the way it protects against indeterminacy is very similar to how Magnetism does it—according to both views the non-linguistic facts determine which interpretation is the correct one.

ii. Biosemantics: The Checklist

We have obtained an affirmative answer to the first question on the metametasemantic task checklist, and we can now consider how Biosemantics handles the remaining requirements.

- **MMS1.** Does it resolve indeterminacy?
 - **YES.** With ease.
- **MMS2.** Does it anchor metasemantics in non-linguistic facts (and is compatible with those facts as well as scientific theories concerning those facts)?
 - **YES.** By grounding metasemantic justification in evolutionary biology, Biosemantics offers a robust framework for explaining why certain meanings and references are preferred.
- **MMS3.** Does it provide tools for comparing metasemantic theories (and a framework for metasemantics that is continuous with them)?

¹² As a matter of fact, we can imagine a counterfactual situation where a magnetist possesses some theory of naturalness. Let's imagine we are dealing with an eccentric magnetist who is also a kind of a nominalistic materialist—in her view, only material objects exist, and naturalness is reduced to spatial boundaries, i.e., each object that has defined spatial boundaries is a separate natural kind and there should be a separate predicate corresponding to that object. On this basis, it would be easy to eliminate the indeterminacy of metasemantics. However, the eccentric magnetist will face similar problems as the biosemanticist—namely, she will struggle with the indeterminacy of the boundaries of objects. As we know, the physical boundary between an apple and the surrounding air is essentially a cloud of electrons and other sub-particles, among which it is impossible to distinguish which belong to the apple and which to the air. Whether this indeterminacy stems from incomplete empirical data about the apple or is an intractable indeterminacy of the language in which the scientific theory is formulated remains an open question. A similar question applies to the selection of proper functions that are supposed to correspond to a given semantic phenomenon. Thus, both the magnetist and the biosemanticist will not be free from indeterminacy at the described level, but this does not prevent them from asserting that one of the object's boundaries is the correct one and that one of the proper functions is the correct one—and therefore, that one of the alternative interpretations of the language is the correct one.

- **YES.** Biosemantics (like Magnetism) only allows for naturalistic theories. Moreover, it is backed by a specific empirical scientific theory, which provides Biosemantics with the tools to decide between alternative naturalistic metasemantics by using the criterion of consistency with evolutionary biology.
- **MMS4.** Does it allow for the application of different metasemantic theories to different linguistic phenomena?
 - **YES.** Biosemantics can accommodate pluralism by recognizing that different types of meanings and references might be explained by different evolutionary functions. This flexibility allows it to handle multiple perspectives within a cohesive framework.

So, as can be seen, the final verdict is that Biosemantics fulfills all four tasks designated for a metametasemantic theory.

5. Magnetism and Biosemantics: Comparison and Conclusion

Biosemantics was designed as a semantic or metasemantic theory, but as has been shown, due to the problem of indeterminacy of proper functions, it has a significant issue in fulfilling the tasks expected of semantics or metasemantics. However, the requirements for metametasemantics are significantly different, and as I have demonstrated, in the key matter of securing metasemantics against indeterminacy, Biosemantics performs no worse than Magnetism.

Ultimately, both accounts succeed as metametasemantic theories by grounding meaning in objective, non-linguistic facts, thereby solving the core problem of radical indeterminacy. They achieve this through different but equally valid strategies. Magnetism appeals to the metaphysical structure of reality, while Biosemantics roots meaning in the functional norms established by evolutionary history. Each demonstrates remarkable explanatory power within its domain. Magnetism elegantly explains why certain referents (like gold rather than gold-or-pyrite) are privileged through microstructural unity, while Biosemantics accounts for the dynamic, historically contingent nature of meaning by tying it to concrete adaptive advantages rather than static metaphysical categories.

The robustness of Biosemantics becomes particularly apparent when examining three key dimensions where it, in my opinion, subtly outperforms Magnetism. First, its functional approach naturally handles cases of biological ambiguity that would challenge naturalness-based accounts. While hearts unambiguously pump blood regardless of taxonomic debates about species/natural kinds boundaries, Magnetism struggles when physical theory presents equally “natural” candidate properties (like mass versus charge in fundamental physics) without specifying which should dominate reference. This difference stems from Biosemantics’ capacity to appeal to the stabilizing effects of selection pressures—minute fitness advantages that accumulate over generations inevitably break apparent symmetries between competing functions. Even in transitional evolutionary phases where traits appear multifunctional, the historical trajectory of adaptation provides principled criteria for meaning determination that Magnetism lacks.

Second, Biosemantics demonstrates superior resilience to theoretical revolutions. The common objection that Biosemantics would collapse if evolutionary theory were falsified misunderstands its core commitment: not to Darwinian mechanisms *per se*, but to the broader principle that functional organization in living systems necessarily generates stable semantic norms. Even radical scientific shifts (to epigenetic, cultural, or yet-unknown inheritance systems) would preserve the essential insight that representational stability emerges from adaptive success. Magnetism lacks this flexibility—without its unspecified “final physics” to ground naturalness, the theory has no fallback position, as natural properties are defined precisely by their role in fundamental laws.

Third, Biosemantics better accommodates the empirically observed plasticity of meaning in natural language. Where Magnetism must posit fixed, language-independent natural kinds, Biosemantics recognizes that representational norms can shift along with changing adaptive pressures—explaining phenomena like semantic change or context-dependent meaning without compromising determinacy. This dynamic grounding proves crucial for handling real-world language use, where meanings remain stable enough for communication yet flexible enough to evolve.

These advantages, while individually subtle, collectively give Biosemantics a marginal but decisive edge. Both theories satisfy the core metametasemantic criterion: they eliminate indeterminacy (MMS1). Moreover, with greater or lesser success, these theories nevertheless meet criteria MMS2 – MMS4.

However, Biosemantics achieves this with greater empirical accountability and theoretical flexibility. Its ability to explain meaning as simultaneously

biologically constrained and historically situated offers a more comprehensive account of semantic phenomena. This is not to dismiss Magnetism's considerable strengths. For domains where functional considerations are irrelevant (like mathematical language), it remains an invaluable tool. Moreover, with refinements—particularly a more developed account of how physical properties translate into referential constraints—it could achieve full parity with Biosemantics. But as currently formulated, Biosemantics' capacity to bridge abstract meaning with concrete biological processes makes it the more robust choice. Its minimal superiority exemplifies what metametasemantics requires: not just a solution to indeterminacy, but one that connects meaning to the real mechanisms that sustain it.

Acknowledgments

I am deeply grateful to both reviewers for their insightful comments, which were instrumental in refining this article and elevating its academic merit. The work on this paper was funded by National Science Center, Poland, under grant number 2018/31/D/HS1/03745.

REFERENCES

- Brandom, Robert. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.
- Brandom, Robert. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.
- Burgess, Alexis, and Brett Sherman, eds. 2014. *Metasemantics: New Essays on the Foundations of Meaning*. Oxford: Oxford University Press.
- Cummins, Robert. 2002. "Neo-Teleology." *In Functions: New Essays in the Philosophy of Psychology and Biology*, edited by André Ariew, Robert Cummins, and Mark Perlman, 164–174. Oxford: Oxford University Press.
- Dretske, Fred. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.
- Field, Hartry. 1980. *Science Without Numbers: A Defense of Nominalism*. Princeton: Princeton University Press.
- Fodor, Jerry. 1990. *A Theory of Content and Other Essays*. Cambridge, Mass: MIT Press.
- Gould, Stephen Jay, and Richard C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the

- Adaptationist Programme.” *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205 (1161): 581–98.
- Kawczyński, Filip. 2021. “In Defence of Metametasemantics.” *Axiomathes* 31: 401–418. <https://doi.org/10.1007/s10516-019-09467-6>
- Kovacs, David. 2020. “Metaphysically Explanatory Unification.” *Philosophical Studies* 177 (6), 1659–83.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Mass: Harvard University Press.
- Ladyman, James, and Don Ross. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
- Lewis, David. 1983. “New Work for a Theory of Universals.” *Australasian Journal of Philosophy* 61 (4): 343–77.
- Lewis, David. 1986. *On the Plurality of Worlds*. Cambridge: Blackwell Publishing.
- Lewis, David. 1984. “Putnam’s Paradox.” *Australasian Journal of Philosophy* 62 (3): 221–36.
- Manley, David, and Robert Wasserman. 2008. “On Linking Dispositions and Conditionals.” *Mind* 117 (465): 59–84.
- Massimi, Michela. 2017. “Laws of Nature, Natural Properties, and the Robustly Best System.” *The Monist* 100 (3): 406–21.
- Miller, Alexander. 2008. “Meaning Skepticism.” In *The Blackwell Guide to the Philosophy of Language*, edited by Michael Devitt and Richard Hanley, 91–113. Malden, Mass.: Wiley-Blackwell.
- Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, Mass.: MIT Press.
- Millikan, Ruth Garrett. 1989. “Biosemantics.” *The Journal of Philosophy* 86 (6): 281–97.
- Millikan, Ruth Garrett. 2000. *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press.
- Neander, Karen. 1991. “The Teleological Notion of ‘Function’.” *Australasian Journal of Philosophy* 69 (4): 454–68.
- Neander, Karen. 2002. “Types of Traits: The Importance of Functional Homologues.” In *Functions: New Essays in the Philosophy of Psychology and Biology*, edited by André Ariew, Robert Cummins, and Mark Perlman, 390–415. Oxford: Oxford University Press.
- Neander, Karen. 2006. “Naturalistic Theories of Reference.” In *The Blackwell Guide to the Philosophy of Language*, edited by Michael Devitt and Richard Hanley, 374–391. Malden, Mass.: Wiley-Blackwell.

- Nencha, Cristina. 2017. "Natural Properties Do Not Support Essentialism in Counterpart Theory." *Argumenta* 2 (2): 281–92.
- Papineau, David. 1987. *Reality and Representation*. Cambridge: Blackwell.
- Papineau, David. 2001. "The Rise of Physicalism." In *Physicalism and Its Discontents*, edited by Carl Gillett and Barry Loewer, 3–36. Cambridge: Cambridge University Press.
- Putnam, Hilary. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Price, Huw. 2011. *Naturalism Without Mirrors*. Oxford: Oxford University Press.
- Schiffer, Stephen. 2003. *The Things We Mean*. Oxford: Oxford University Press.
- Schaffer, Jonathan. 2009. "On What Grounds What." In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David J. Chalmers, David Manley, and Ryan Wasserman, 347–83. Oxford: Oxford University Press.
- Schwarz, Wolfgang. 2014. "Against Magnetism." *Australasian Journal of Philosophy* 92 (1): 17–36.
- Sider, Theodore. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.
- Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
- Simchen, Ori. 2017. *Semantics, Metasemantics, Aboutness*. Oxford: Oxford University Press.
- Soames, Scott. 2005. *Reference and Description: The Case Against Two-Dimensionalism*. Princeton: Princeton University Press.
- Stich, Stephen. 1992. "What Is a Theory of Mental Representation?" In *Mind* 101 (402): 243–61.
- Weatherson, Brian. 2013. "The Role of Naturalness in Lewis's Theory of Meaning." *Journal for the History of Analytical Philosophy* 1 (10), 1–18.
- Williams, J. Robert G. 2007. "Eligibility and Inscrutability." *Philosophical Review* 116 (3): 361–99.
- Wilson, Mark. 1982. "Predicate Meets Property." *The Philosophical Review* 91 (4): 549–89.
- Wright, Crispin. 1992. *Truth and Objectivity*. Cambridge, Mass. Harvard University Press.