

UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Studi Umanistici

Corso di Laurea Magistrale in Scienze Filosofiche



NATURALIZING KRIPKENSTEIN: HOW
PRIMITIVIST, DISPOSITIONAL AND SKEPTICAL
ANSWERS TO KRIPKE'S WITTGENSTEIN ALL FIT
WITHIN AN EVOLUTIONARY ACCOUNT OF
MEANING

Relatore:

Prof. Andrea GUARDO

Correlatore:

Prof. Paolo SPINICCI

Tesi di Laurea di:

Dario VACCARO

Matr. n. 938603

Anno Accademico 2019/2020

Section 1: Setting the stage

In 1982, Saul Kripke published a book meant as an unusual and revolutionary analysis of Wittgenstein's later philosophy of language¹ but destined to become the foundation for much more: in the almost 40 years since its introduction in the literature, *Wittgenstein on Rules and Private Language* (WRPL) has yielded countless replies, comments, interpretations and further ideas in diverse areas of philosophy, including epistemology, metaphysics, social philosophy, and more. To date, no consensus has been reached on its fundamental claim, which will be the focus of this work: the idea that we mean nothing by our words.

In Kripke's own terms, his work was supposed to constitute "an elementary exposition of what [Kripke takes] to be the central thread of Wittgenstein's later work on the philosophy of language" (p. VII), that is, the problem of rule-following. We can find an explicit statement of it in Wittgenstein (1953)'s famous §201: "This was our paradox: no course of action could be determined by a rule, because any course of action can be made out to accord with the rule". In retrospect, it is quite hard to see how Kripke's work could be understood as "an elementary exposition"; after all, Kripke himself, in the same Introduction where he describes WRPL in such way, seems quite unsure about the adherence of his interpretation to what Wittgenstein really wanted to say². To be sure, WRPL is now generally considered an inaccurate understanding of what Wittgenstein claimed: in particular, the preeminence of the "congruence of behavior" over "language norms" (Sládeček 2020) seems to be a misattribution, as I find compelling due to Wittgensteinian passages such as (1953), §185.

But even if WRPL may have historically failed in exegetical terms, it is now almost universally regarded as one of the greatest challenges to the intuitive picture we have of a number of topics in the philosophy of mind and the metaphysics of meaning, going as far as, at least in the latter case, shattering the very foundations the discipline is normally thought to be built on: meanings and their connection to our utterances. Precisely because most commentators deny that ascribing to Wittgenstein these ideas would be a correct interpretative move, they have been ascribed to a fictional character by the name of *Kripenstein*; such name is intended as implying that this figure is Wittgenstein as *seen*

¹ Mainly found in the *Philosophical Investigations* (1953) and in the *Remarks on the Foundations of Mathematics* (1956).

² As confirmation, in various moments of WRPL Kripke makes it very clear that his exposition may not be totally adherent to Wittgenstein's intentions ("Probably many of my formulations [...] are done in a way Wittgenstein would not himself approve", WRPL p. 5). Not everyone in the literature agrees with him, though, claiming that this *is* the correct reading of the late Wittgenstein. See, for instance, Kusch (2006).

through Kripke's perspective³. If we assume, as we will, that WRPL really does not represent a loyal interpretation of Wittgenstein's later philosophy⁴, this has an immediate exegetic implication: it tells us that whatever claims KW fictionally supports, they are not necessarily supported by either Kripke himself or by the real Wittgenstein. As Kripke famously states, "The present paper should be thought of as expounding neither 'Wittgenstein's' argument nor 'Kripke's': rather, Wittgenstein's as it struck Kripke, as it presented a problem for him" (WRPL, p. 5). Of course this can represent a problem for historians of philosophy who try to assess where KW, Wittgenstein and Kripke's positions collide and where they agree; we will not need to be particularly worried about this, since this work is entirely focused on KW's argument itself, under a purely theoretical perspective and regardless of the history of its literature.

That being said, at least some of the arguments *against* some replies to KW are clearly endorsed by Kripke himself, which is normally signaled by the author's usage of the first person. In other instances, Kripke clearly speaks against KW and the claims the fictional figure endorses. In such sorts of cases, I will refer to the real-life author as the proponent of the objections or as the critic of KW's claims; when it is not clear who supports a claim or theory, I will instead ascribe them to KW, since "he" is, by Kripke's own acknowledgement, the main representative of what is argued in WRPL.

Since its publication, WRPL has seen countless replies and comments from hundreds of professional philosophers⁵. Most of them, with the dispositionalists in the forefront, side with common sense by rejecting KW's fundamental thesis, which I will introduce and explain in this Sections, for readers unfamiliar with it. KW claims that there is no such thing as "meaning something" by our words, signs, utterances: when I call my dog by his name, or write down a list for grocery shopping, the content of my words mean nothing.

Initially, we should not be surprised by the fact that the general consensus finds KW's claim to be false: it may actually look more surprising that anyone ever agreed with him. As we will shortly see, KW's claim is usually (and understandably) seen as a *threat* to the natural understanding of our rational practices, in a similar fashion to the classic, and much older, skeptical challenges about our knowledge of the world. Similarly to the latter philosophical argument, in fact, its extreme dissonance with our naïve intuitions and its, at

³ For brevity, I will refer to Kripkenstein as "KW" (Kripke's Wittgenstein). In WRPL, Kripke usually refers to the fictional philosopher as "the sceptic": this nomenclature is not necessarily unacceptable, but it may be too generic and implies that KW's claim really is skeptical, which I find to be a controversial position.

⁴ For a more detailed defense of this point, which will not be extensively explored in this work, see McDowell (1998).

⁵ It can be estimated, on the basis of Google Scholar data, that the scholarly articles available on the internet that quote "Kripkenstein" are at least six hundred. This is not counting the works where the problem is addressed, but the word is not used.

least apparent, practical inapplicability make it one of those theories one reads about or learns about in classes, but that could never become an actual part of one's set of beliefs.

As such, this essay's acceptance of KW's basic claim as a correct philosophical thesis seems to put a great amount of pressure on the philosopher in order to make its content acceptable to their public and themselves. After all, it is not unusual for philosophers to tend *a priori* towards a defense of our intuitions, especially when they are extremely widespread, such as the idea that words have meaning and that we mean something by our words when we speak. There is nothing wrong with that, of course: if we can preserve an idea that most people already have, that can be considered, at the very least, as epistemic savings. It should be equally clear, though, that philosophers also need to revise our common intuitions when they are found to be untenable. I will try to show that this is the case.

Following what we just said, there are some authors who accept KW's conclusion and try to make sense of it, usually by making (more or less) slight changes to his overall view, that, with no exception, aim at making KW's claim more digestible. This essay is intended as one such attempt: I support the idea that KW's "problem" needs no real solution, as it is just a correct statement of how things are, and I will try to show why this should in no way preoccupy us; once properly understood, the "naturalistic" solution should turn out to be quite intuitive itself. Moreover, the wider goal I will be aiming at is to show how the main competing answers to KW's problem fit⁶ within the framework that justifies this so-called "skeptical solution"⁷, thus making it clear that KW's account of meaning is not a challenge to the notion of meaning itself, but rather the best way to make sense of it in a naturalistic framework, despite what KW or Kripke themselves may think.

Now that the cards are on the table, it seems important to better understand both KW and the famous replies I will be taking on.

KW's claim

KW states that "There can be no fact as to what I mean by [...] any [...] word, at any time"⁸. For now, let us call this "The Paradox", in line with Kripke's terminology. It seems clear why this statement would have the appearance of a paradox: KW gives reasons to think that meanings do not exist, but if there is no fact of my meaning anything by any word, *what do I mean by that same statement?* It seems we are in the presence of what Mackie (1964) calls "pragmatic self-refutation"⁹: KW introduces an idea in the form of a written

⁶ By this I mean that the competing solutions are wrong as replies to KW, but they nevertheless grasp an aspect of communication.

⁷ In the next few pages, I will discuss more in depth this problematic nomenclature.

⁸ Kripke (1982), p. 21.

⁹ "We have *pragmatic* self-refutation when the way in which something is presented conflicts with what is presented" (p. 193).

text, which is understandable by the reader, supposedly, because of its meaning. But the *meaning* of this line of text is that lines of text (alongside spoken words, signs, etc.) *have no meaning*. It seems clear that, if KW is right, we are facing a powerful philosophical dilemma.

In order to analyze the Paradox more straightforwardly and to outline it more precisely, I will address it by answering three basic questions. I believe this is essential, given that KW's problem is so insidious and difficult to formulate that it sometimes takes longer to understand it rather than to accept one of the answers to it.

1. *What does KW refer to exactly?*

To start understanding the Paradox better, we need to make a clear distinction between two closely related homonyms in the English language, which seem to me to be the cause of many misunderstandings in the Anglophone literature on this topic:

1. "Meaning" as in "the meaning of a word": the concepts we -in a more or less metaphorical sense- *attach* to words and the Platonists believe exist outside our linguistic practices;
2. "Meaning" as in "by the word *x*, they meant *y*": the intentional act of attaching a concept to a word.

Since the English language doesn't seem to have a sufficiently diverse vocabulary in this field, making it hard to make this distinction clearly enough, I will call the first *meaning-S* and the second *meaning-I*¹⁰.

With this distinction in mind, we can already achieve a clearer understanding of the Paradox, at least in its *initial form*, which we will rephrase in this way:

"There are no intentional acts of the *meaning-I* kind".

The rephrasing we just adopted shows us something else that is worth pointing out. KW argues for the non-existence of *meaning-I* acts; he does not say anything about the existence of intentional acts in general. In philosophy of mind, intentional acts are commonly understood as mental states characterized by the property of being *about* something or *directed to* something. *Meaning-I, if it exists*, is an intentional act: to mean something by a word is an act of addressing something, of attaching a *meaning-S* to it. But

¹⁰ "S" stands for "sense" and "I" stands for "intentional act". The distinction will be omitted whenever the context or the phrasing makes it obvious which one I'm referring to. In other contexts, I will use the expression "meanings" when both *meaning-I* and *meaning-S* are involved. Using the verb "to intend" instead of "to mean" is also a possible alternative, but I fear it would be just as confusing, given the prominence of terms such as "intentionality" and "intentional acts" in the literature, which are related but not identical to it.

there are many other intentional acts, such as perception, desire, fear. KW is *only* stating that meaning-I does not exist. All other intentional acts thus seem safe from his attacks¹¹.

With these precautions, a large part of the misunderstandings in the literature on KW have already been addressed: I will return to them when the mix-ups will come up in some of the arguments I will analyze.

2. *What does the Paradox imply?*

The claim that there is no meaning-I implies that whenever one speaks, writes, signals something, there is in principle no telling what they want the other person to understand. When I say “no telling” I mean it literally: KW doesn’t just say that we have no way to know which meaning-I our interlocutor picked; he claims that there is nothing to know in the first place, since there is no such thing as meaning-I. Let us see this with an example. At a greengrocer’s, I ask “May I have an apple?”. Now, a skeptical seller could ask herself: “What if by ‘apple’ he means what I would call ‘banana’?”. This is a legitimate question, and the starting point of KW’s argument¹². But his Paradox goes further than that, because the conclusion is that neither I nor the seller actually mean *anything* by “apple”.

Of course, especially after contextualizing it in a familiar setting, this claim sounds perfectly insane¹³.

3. *Why exactly does KW come up with this – at least prima facie- insane idea?*

The Paradox arises in KW’s “mind” (remember we are talking about a *fictional* philosopher) when he asks himself: “How do I know what x means by y ?”, that is, the epistemological question. The answer to this question would be expected to be a matter of fact that determines my knowledge: I know that my shirt is green because of a set of facts (its pigmentation, the reliable intentional act of seeing it...), so there must be something “out there” that grounds semantic facts, justifies my beliefs about them and turns it into knowledge¹⁴.

Unfortunately, throughout WRPL, KW shows that every candidate for the role either doesn’t exist or is not an appropriate justification. By elimination, he concludes that the most reasonable answer is that the desired matter of fact just isn’t there. Since KW believes

¹¹ Unless other intentional acts were grounded in meaning-I, but this is never hinted at by KW, at least explicitly. Boghossian (1989) does claim that the challenge posed by KW extends to *all* intentional acts: I will address his related arguments in Section 2. Chalmers (1996) claims that “[KW] argues in effect that there can be no entailment from physical and phenomenal facts to intentional facts”, but in his case no explanation is given for such an interpretation: I thus find plausible to rule out his claim as a misunderstanding of WRPL.

¹² The “epistemological question” (“How do I know what x means?”) may actually be more closely represented by Quine’s “indeterminacy of translation” than by KW’s Paradox.

¹³ As Kripke himself famously writes: “if the sceptic proposes his hypothesis sincerely, he is crazy” (WRPL, p.9) and “the sceptical conclusion is insane and intolerable” (WRPL, p. 60).

¹⁴ Gettier cases notwithstanding.

that the epistemological question can have an answer only if there is a positive answer to the metaphysical question (that is, the matter of fact we need to justify our belief must actually exist), then he *also* concludes that we can never know what anyone means by anything.

A first reasonable objection to KW's approach could appear to be that, given the alleged absurdity of the skeptical conclusion, it is far more likely that he missed something over the course of his analysis than that the conclusion be actually right. In other words, some may think that we should infer from the paradoxical nature of the conclusion KW comes to, that his argument is wrong. Specifically, even if we were to find every argument KW uses to discredit the meaning determinist accounts he rejects as sound, there are at least two different routes he could have taken instead of accepting the defeat of meaning determinism as a whole: one is to suppose that he has not considered *every* plausible account of meaning-I; the other is to suppose that his overall approach to meaning determinist accounts is fundamentally flawed or that his understanding of said accounts is questionable. Most of the commentators that disagree with KW's conclusion have followed one of the two routes; the most common example is found in the literature defending dispositionalism: some argue that KW has correctly refuted the versions he discusses, but that he has missed at least another version of dispositionalism that doesn't fall prey to his attacks (see, for instance, Heil & Martin's (1998) use of "finkish" and masked dispositions), while others believe that his general understanding of dispositionalism is too narrow and/or oversimplified (for instance, Warren 2020).

My take on this -and the take of most contemporary authors who support the skeptical answer- is that the problem isn't in the paradoxical nature of the conclusion: the problem is to *think* that KW's conclusion has a paradoxical nature, something Kripke seems to take as a given. That is why the Paradox still lives on as an actual answer to the skeptical question about meaning-I instead of just being part of a *reductio* argument. This is also why, instead of using expressions that suggest internal incoherence, I believe it is good to give KW's conclusion a different name.

Martin Kusch, in his seminal (2006), gives an excellent outline of the competing perspectives on KW's challenge. He calls those who oppose KW's conclusion "meaning determinists"¹⁵, implying that the fixed meaning-I of a word *determines* the way we will and/or ought to use it. He then calls the defenders of KW's conclusion "meaning skeptics", arguably choosing this wording because of the famous expression used in WRPL, "skeptical answer".

The problem with this terminology is that "skepticism" has come to denote in the contemporary literature two very different perspectives, depending on the philosophical field: one where there is a matter of fact in the world, but we have no epistemic access to it;

¹⁵ Kusch (2006), p. 4.

the other, and this is KW's case, where there is *no* such matter of fact and *because of that* we have no justification for a belief that is grounded in it. Thus, since I fear that talk of "skepticism" in this context could lead to misconceptions, I will call the supporters of KW's conclusion "meaning indeterminists". The name refers to the claim that there are no meaning-I facts at all and that, because of that, meaning-S is always indeterminate when we communicate; it is also a more natural way to refer to the deniers of the meaning determinist claim.

Let us now look further into this idea, in order to frame our problem in a more definite way. KW asks us to think what the meaning attached to a word, such as "chair"¹⁶, is, when we utter it, write it down, read it. Of course, our intuitive idea is that "chair" just refers to chairs in the world (actual or possible). But what if -and here comes the properly "skeptical" thought process- "chair" had had the following meaning all along:

Chair-2030:

Until the year 2030: refers to chairs

After the year 2030: refers to tables

Once again, this does sound insane. But the real problem is that there seems to be no way to explain why that would *in principle* be a wrong application of the term. Specifically, KW argues, there is absolutely no difference (at least until the year 2030) between this use of "chair" and the one we would intuitively expect, either internally, in our psychological life, or externally, in our behavior. It can be easily seen that this is a problem concerning rule-following in general: we could similarly say that people playing chess are actually following the "Chess-2030" rule, such that from 2030 on they will have to capture the Queen in order to win, or something like that. Thus, there is no way to determine whether we have always followed the intuitive rule, the "Chair-2030" one, or *any* of the infinite other rules that are compatible with the facts about our previous usage of the word "chair". In conclusion, KW effectively argues that whenever we assume we mean x by our words instead of y (or one of the infinitely many other alternatives) we do so in a *completely arbitrary way*, as long as both x and y (or any of the other alternatives) are rules compatible with our previous behavior. Though I agree with the overall argument, I will be challenging this last assumption.

A field distinction¹⁷

¹⁶ The original example is the notorious addition/quaddition problem, but since it is more problematic and harder to grasp, it seems better to introduce the problem with a simpler one. I also make abundant use of Goodman's riddle of induction because it can be applied to any word whatsoever. Moreover, as Guardo (2019) points out, "[Goodman examples] immediately make clear that Kripkenstein's Paradox applies also to the cases the speaker has already dealt with". That said, the addition/quaddition example, when properly understood, is perfectly analogous and does not imply any loss of generality.

¹⁷ I am widely indebted to Guardo (in his 2020a and from private conversation) for this section.

Another widespread confusion in the debate is the philosophical field KW is interested in. Is this a problem in the philosophy of mind? Or is it, rather, a metaphysical question? It seems, once again, that identifying KW's question may be almost as hard as answering it, and, in this case, WRPL can help us only to an extent.

We have already established that the fundamental question on KW's mind is *not* epistemological¹⁸: it is not merely a matter of how I know what I mean by a word, but of what makes it the case that I do. Importantly, this is also not a matter of the existence of meanings-S: it is not a debate on whether the existence of meanings ground our use of them or the other way around. In other words, one could coherently take a determinist or indeterminist position in this debate whether they believe meanings-S to exist (mind-independently) or not. That is because KW wants to discuss the *connection* our signs have with meaning-S, whatever its ontological status may be. That connection is what we have come to call meaning-I, an intentional state. As such, at least one part of the problem *is* correctly ascribable to the philosophy of mind: KW questions the nature and the existence of a specific mental state, the act of meaning something by a word. This first question will sound familiar by now: "Is there a meaning-I intentional state? And, if there is, what makes it distinguishable from other intentional states?". KW's answer is that there is no such state.

Even though that is the starting point of KW's investigation, it is still not all there is to it. Once KW has given sufficient evidence for the non-existence of meaning-I, in fact, a different question becomes even more prominent over time: a question in the metaphysics of meaning, or metasemantics; that is, a question about meaning-S. Our new question can be framed in this way: "What determines the meaning of a word?". This is a matter that does not *immediately* concern our mental states: rather, it questions the relation between a signal and what it is supposed to signal. Notice the subtle difference with the first question: the first is a problem about the connection between a mental state and what rule it is supposed to conform to; the second is a problem about what this rule really consists of, and how it comes to be. In this second question, the object of analysis turns out to be meaning-S, the very concept Kripke warned us about: we may very well argue that the entire "solution" KW outlines in the second half of WRPL is an attempt to solve this problem. But, the reader may recall, we said that we are in no way concerned with the *existence* of meaning-S: that remains true, because the second question asks us in virtue of what meaning-S is what it is, not whether it really exists or not. This has two theoretical justifications: first, it doesn't require us to have a strong ontological commitment in favor or against meaning-S; second, it helps us with coming to terms with the intuitive idea that we *have to* make sense of meaning, even if maybe not in the strong terms of meaning determinism.

¹⁸ WRPL, p.21.

Thus, the second question is about the nature of meaning-S, not about its existence. KW's answer to this problem is less definitive than the first one, but it seems to be connected with the idea that a group of individuals will try to cooperate even though there is always, in principle, the risk of their cooperation not working out, precisely because no *rule* is genuinely attached to their utterances and to the reference of their words. Essentially, meaning turns out to be, in his view, a social agreement that determines the assertability conditions of words in specific contexts. This makes KW's understanding of meaning-S a lot weaker than most, so much so that I would call it a "nominalistic" account, by assimilation with the classic debate on abstract objects.

In conclusion, we could say that the two questions KW challenges us with are two perspectives on a same (supposed) fact: the first-person perspective, where the individuals ask themselves what happens in their mental life when they mean something by a word; the third-person perspective, where a (bizarre) passer-by¹⁹ asks himself what makes it the case that two people he's watching are referring to something specific in the world by the words they use in a dialog. One is about a fact restricted to the mind, the other a fact extended to the world. Notice, though, that both questions are fundamentally metaphysical; they both question the existence and/or the nature of two facts in the world: the question in the philosophy of mind is a question about the existence of meaning-I facts, whereas the question in metasemantics is a question about the nature of meaning-S facts.

The two questions have been mixed-up and confused profusely over the last few decades. Because of the importance of this distinction, it becomes important to give a proper exegesis of each work concerned with answering KW's challenge. A common indicator of what is actually being investigated is the way the question is framed: a common phrasing for the question in the philosophy of mind is to ask "What fact *about me/about my mental life* makes it the case that I mean *x* by my words?"; an equivalent in the metaphysics of meaning would be "What fact makes it the case that a word refers to an object/we are justified in using a word to refer to a certain object?". Of course, since this distinction is not recognized by many writers on the topic, the distinction in some of the works in the literature is not so easy to outline, while some other times answering one of the two questions is seen as a response to the whole problem. For now, let it suffice to say that the dispositional answer is usually looking at the metasemantic question, whereas the primitivist answer usually starts from the problem in the philosophy of mind.

Moreover, saying that the questions should not be mixed together and that answering one is insufficient to shut off KW's worries once and for all, is not equivalent to say that the two questions are *unrelated*. On the contrary, under some aspects answering one amounts to answering the other, but this is not *always* the case, and some arguments we will discuss are strong or useful only for one of them and weaker or irrelevant for the other. In

¹⁹ Of course, a party to the communication can also "watch" the scene he's in as if an outsider.

particular, it is important to notice that even if we deny the existence of any sort of “special” mental state of the meaning-I kind, we can (and I will) accept the existence of communication (i.e., a ground for semantics) in behavioral terms. My main focus will be on the second question, because it *needs* explaining: the fact that we do communicate successfully all the time is evidence enough of this need; but discussion of the problem in the philosophy of mind is essential as well, because of the close relation the two aspects share, especially if a stronger sense of meaning-S is being defended. It will be, of course, of great importance to understand whether one of the two aspects grounds the other or supervenes on the other: in particular, in order to support the claim that communication exists but meaning-I doesn’t, I will need to show that it is not the case that reference is grounded in the intentional act of meaning-I, as some²⁰ philosophers think, because if that were true either both exist or neither does.

One last metaphysical distinction

To conclude our introductory analysis of KW’s riddle, I would like to stress that there is a further distinction that needs to be settled by any theory attempting to solve it. The philosopher answering to KW needs to make his metaphysical stance clear: are we going to be realists, non-realists or, as was recently suggested²¹, *quasi*-realists about meaning-I? And what about meaning-S? Let us clear up the terminology, since it seems very important to understand why people may have different things in mind when they say that “Paul means x by y ”, and because it will come up and become prominent in Section 3. Further misunderstandings in the literature will thus become apparent.

Realism about meaning-I: This is the claim that the existence of meaning-I is a fact in the world, i.e., that in the list of things that exist there is also meaning-I. It is also the claim that whenever we say “Carl means x by y ”, the truth-value of the proposition associated to this statement depends on the occurrence of said states of affairs independently of the semantic and mental facts about the proposition itself, in the same way that we accept “The snow is white” to be true iff the snow is white in the world.

Anti-realism about meaning-I: This is the claim that meaning-I is not among the things that exist mind-independently²². When we say that “Carl means x by y ”, we are saying

²⁰ As with most philosophical issues, it is not easy to assess the general consensus on this topic. That said, the most recent attempts to ground meaning have focused, among others, on “intentionality” (Shafiei 2018), “[Donald] Davidson’s triangulation” (Verheggen 2019) and “intersubjectivity” (Sandler forth.). Thus, the consensus seems to be, at least, that meaning is grounded in phenomenal and/or behavioral facts of a community of speakers.

²¹ By Miller (2020).

²² I believe that this specification is needed since Schaffer’s (2009) claim that “everything exists”, in the meta-metaphysical context of grounding relations. Of course, there is a sense in which meaning-I is mind dependent: namely, due to the fact that it is a *mental* state. In this sense, there would be no meaning-I without a mind to “produce” it. What is in question here is different: we are questioning whether its

something strictly false, because there is no meaning anything by any word. This *may* be what KW has in mind and it is definitely the standard understanding of his position, due to assertions such as the fundamental one we started with: “There can be no fact as to what I mean by any word at any time”.

*Quasi-realism*²³ about meaning-I: This is the claim that meaning-I does not really exist mind-independently, but also that we need or ought to act as if it were. Clearly, this is still an anti-realist claim: in this view, meaning-I still does not *really* exist. That said, meaning-I is given an existence in fictional terms: we make-believe that it exists in the same way that we make-believe a dragon exists *for the purposes of a game*. On this view, the existence of meaning-I is helpful in communication, due to it allowing for a number of *language* games that would be otherwise impossible, or hard to follow. In this case, when we say that “Carl means *x* by *y*”, we are saying something meaningful but not *strictly* true, or not true to reality: one problem this position needs to face is to show that it is sometimes permissible, or even better than the alternative, to believe something false²⁴.

Exegetically speaking, this may also be what KW has in mind, but, even if it wasn't, it could still represent a good way to frame the meaning indeterminist answer, as I will suggest.

Whatever the correct answer, then, it seems that we have two new problems to deal with in addressing KW's challenge:

(1): Is meaning-I a (mind-independent) fact in the world? If so, what, if anything, grounds it? Is it meaning-S? Some non-semantic fact about the world or about our minds?

(2): Whatever the answer to (1), can we make sense of meaning-I attributions and assign them a precise rule for correctness? That is, if we accept that meaning-I is there, do we have epistemic access to its nature?

The main opponent: dispositionalism

existence is the product of our imagination or not. In this sense, we can similarly say that a unicorn exists, *as the product of our imagination*.

²³ “Quasi-realism” is a term coined by Blackburn (1984a) as a means to explain our attitude towards moral “facts”. In Blackburn's view, moral facts don't strictly exist outside our minds, but we talk about them as if they did. The similarity is striking: thus, it makes sense for Miller to have reused the term in the metasemantic context.

²⁴ Though it seems that the wide majority of philosophers would agree that one is better off believing some false propositions about the world, one recent opposition to this claim is found in Greenberg (2020). Hattiangadi (2019) gives a direct reply. Naturally, this discussion is relevant if we accept the (widely held) correspondence theory of truth: given that, as the quasi-realist agrees, meaning-I does not exist, then it is false (not corresponding to facts) that we mean something by our words, and yet the quasi-realist claims that we still should. If the quasi-realist follows a deflationary theory of truth, though, then they can unproblematically claim the truth of “I mean *x* by *y*” even if they support the non-existence of meaning-I. Importantly, this last position has been attributed to KW by, among others, Guardo (2020b).

The most popular strategy to defend meaning determinism has historically been, and continues to be, dispositionalism. In this section we will only introduce the elementary ideas that constitute the main varieties found in WRPL, without discussing them in depth or attacking their weaknesses outside of what KW points out, which will be the focus of Chapter 2.

In WRPL, Kripke puts the basic idea of dispositionalism about meaning in these terms²⁵: I simply mean to refer to chairs when I use the word “chair” because I would be disposed to say “chair” *whenever* I want to refer to chairs; I would be following the *Chair-2030* rule if, and only if, I were disposed to say “chair” to refer to chairs before 2030 and to tables from 2030 onwards. In other words, dispositionalism tells us that it does not matter that our *previous* behavior is compatible with both *Chair* and *Chair-2030* rules and, thus, indeterminate: what matters is that in a future context I will still intend to refer to chairs in the same way I did now, and that is why we *mean* chairs when we say “chairs” and we could not mean anything else. The facts “about us” that determine our meanings are conditional, rather than factual, but they are there. This being the simplest dispositional account we will see, it is labelled the “Simple Dispositional Answer” (SDA)²⁶.

Kripke (in this case aligned with KW) famously points out three main objections to SDA: the problem of finitude, the problem of error, and the problem of normativity. *Finitude* is presented by Kripke as the claim that “not only my actual performance, but also the totality of my dispositions, is finite”²⁷: there are cases, even in the hypothetical form of dispositions, where it is not clear what I would be disposed to do or say or where it is not clear that I would have any disposition at all. *Error* argues that sometimes we are disposed to use words in ways we don’t actually *mean*²⁸. Say you know two twins, Mark and Paul, and one day you mistakenly call “Mark” Paul or vice versa. Your being disposed to make this mistake in certain situations shouldn’t change the fact that the meaning you ascribe to that name is incorrect. Importantly, that is also true from the speaker’s perspective: they are ready to say, “I didn’t mean that”, when they realize they confused Paul with Mark. Thus, intuitively, it seems that there is no direct connection between a speaker’s disposition to use a word and the correctness of its meaning.

*Normativity*²⁹ is often considered the most powerful argument against dispositional answers in general, as it isn’t just a counterexample, but instead it finds dispositionalists to be completely missing the point: “A candidate for what constitutes the state of my

²⁵ Adapted to my example of Chair-2030.

²⁶ I’m following Guardo (2018).

²⁷ WRPL, p. 26. In Section 2 we will see, following Blackburn (1984b), that this formulation is defective, but also that Kripke gives a better one later on.

²⁸ Assuming the dispositional premise that the act of meaning-I exists.

²⁹ *Normativity* is also the most discussed and variously interpreted objection in WRPL: in Section 2 I present the different readings of it and try to assess their strength. For now, I give *Normativity* its most straightforward, or “naïve”, understanding.

meaning one function, rather than another, by a given function sign, ought to be such that, whatever in fact I (am disposed to) do, there is a unique thing that I *should* do”³⁰. And, of course, dispositions cannot be that candidate, because they are a description of what the speaker *would* do in a determinate situation, not a prescription of what they *should* do.

As the name suggests, the SDA is the most naïve version of dispositionalism and, as such, the easiest to tear down. The other version of dispositionalism Kripke considers is the Idealized Dispositional Answer (IDA), which adds a crucial clause to SDA: *x* means *y* by *z* because they would be disposed to follow the rule of *y* in every hypothetical situation, *under ideal conditions*. This seems to prevent at least *Finitude* and *Error* from being valid objections to dispositionalism, because we are asked to assume by definition that there can be no environment-induced mistake or impossibility to attach a meaning to the object in question. Kripke seems to disagree, though. He argues that it is not quite clear what would make it the case that we never make mistakes and that we always have a way to answer the query, concluding that there is no telling how such changes would affect our disposition to answer: “How in the world can I tell what would happen if my brain were stuffed with extra brain matter, or if my life were prolonged by some magic elixir? [...] We have no idea what the results of such experiments would be”³¹.

Though this remark has been seen as problematic and not particularly convincing by some, myself included, it is not the only rebuttal Kripke has in mind against IDA. Actually, Kripke is well aware that the dispositionalist has a (maybe too) simple answer to his objection: “If I somehow were to be given the means to carry out my intentions with respect to numbers that presently are too long for me to add [for instance], I would respond with their sum” (WRPL, p. 28). But in this rebuttal there is, he claims, a clear vicious circularity in asking for “ideal conditions” that agree with my present intentions to mean something by a word. The problem seems to be that the whole point KW is making is that there is no telling whether I intend to use one or any other rule that accords with my previous usage of a word. How, then, can we determine ideal conditions that make us always answer “correctly”, if we have no way to know which rule we were following in the first place?

Although these objections look very promising for discarding dispositionalism completely, many answers have been given to them and dispositionalism is currently as strong as ever under a popularity perspective, with many still giving new insight on where the key to victory for this view may be (the most recent attempts being Båve 2019 and Warren 2020). If it is to succeed, though, there seems to be no way around reworking completely the notion of descriptivity Kripke attacks so vehemently, because any standard

³⁰ WRPL, p. 24.

³¹ WRPL, p. 27.

account of dispositionalism is otherwise doomed to fail in the face of the *Normativity* argument at least.

Primitivism

Primitivism is the other meaning determinist position I will focus on over the course of the essay, mainly due to its being, in my view, the other most interesting, and possibly effective, meaning determinist answer. The essential component in the primitivist strategy is to defend the idea that meaning-I and/or meaning-S are fundamental, that is, not reducible to anything else, similarly to the very plausible fundamentality of some physical facts and the more debated fundamentality of some phenomenal facts³². Defenders of this position are usually convinced by KW's attacks, thus accepting that none of the current theories of meaning are able to get their heads around its nature, but don't find his conclusion acceptable: because of that, the postulation of meaning-I *as a given* is seen as necessary. Kripke gives this position a brief overview and a couple of pages of consideration in WRPL: "Perhaps [meaning-I] is simply a primitive state, not to be assimilated to sensations [...] but a state of a unique kind of its own"³³.

Although he rejects it, Kripke doesn't give us a proper confutation as he did for dispositionalism: primitivism just seems too convenient, as it simply postulates something that it can't explain otherwise. Now, in a sense, this *is* a confutation, by the plausible philosophical rule that one ought not to support a theory without sufficient evidence or arguments to back it up. At the same time, the theory, as it is presented by Kripke, is, in another sense, "irrefutable": to argue against it is similar to arguing that lines should not exist within Euclidian geometry. Then again, this strategy doesn't satisfy very well our philosophical desire for understanding: keeping up the simile, of course we cannot question the existence of lines within Euclidian geometry, since that postulate is intrinsically interlaced with the system itself, but we *can* question the effectiveness of a theory that postulates lines within geometry in general. Thus, I take it that when Kripke calls primitivism irrefutable he has something of this kind in mind, which is also why he feels confident he can reject it.

It is also worth pointing out, while introducing one such theory, that we should not assume every aspect of reality to be explainable or understandable from our point of view: it might be the case that meaning-I exists, but we just cannot get our heads around it, in the same way some Christian philosophers accept that the nature of God is unintelligible, though it is there. If we assume a state of affairs where this image is true, the primitivist has no obligation to give us further details about this "mysterious" fundamental entity, for

³² This analogy will take center stage in Section 3.

³³ WRPL, p. 51.

the simple reason that they can't³⁴. Now, this introduction may have given more reasons to find the theory unattractive than the opposite. But the reason why I find primitivism to have a future in the theoretical discussion is that the recent literature in its support is attempting to give *metaphysical* plausibility to the so-called "primacy of meaning". This is important: instead of claiming meaning to be irreducible due to our epistemic shortcomings, some contemporary primitivist theories attempt to give credibility to the idea that meaning may be a fundamental component of reality itself. Thus, the most promising primitivist accounts in recent years have been those that employ notions of metaphysical "building relations"³⁵, especially supervenience and grounding³⁶: if it can be shown that meaning-I is a fundamental constituent of reality, then there is simply no way to go "further back" and explain it in terms of something else.

It will also be of great importance to define more precisely what aspect of meaning primitivists argue to have a fundamental character. The first, by now obvious, distinction must be made between meaning-S and meaning-I, but among the other aspects we will need to investigate are: how exactly does the metaphysical fundamentality condition us on an epistemological level³⁷; whether this fundamentality holds across possible worlds or if it is contingent to the actual one; finally, whether this fundamentality necessarily links itself to our communication practices, even if it is there.

With all its *prima facie* limitations, it feels apparent to me why the primitivist answer has strength and can look very promising to some, as it did for me when I learnt of the paradox: since KW has laid waste of any plausible explanation of meaning-I, and since meaning-I *clearly exists* -the primitivist argues-, all we can do is accept its existence as presupposed for us to even be able to discuss it or question it. But this is also its downfall: since it says so little of the problem, *any* plausible account that is not annihilated by KW or that gives a satisfying account of meaning-I, will be preferable. I will support the idea that such account exists; thus, primitivism will have to be discarded.

KW's solution

In the last chapter of WRPL, Kripke gives an account of communication he takes, again, from his understanding of Wittgenstein. As such, we will refer to it as "KW's solution" to his own problem. We will see, in Section 4, that I don't take this solution as a viable one, essentially for two reasons: it gives too much power to the community of speakers and too little to the individual speaker; it does not give a convincing reason for the objective

³⁴ Hattiangadi (2019) disagrees. She finds that there is nothing "mysterious" about primitive meaning, and that it can be explained both in metaphysical *and* in epistemological terms. I think there is a misunderstanding here, and I will address it when presenting her position.

³⁵ As Bennett (2011) calls them.

³⁶ For especially good introductions to this relatively new technical notion in metametaphysics, see Fine (2012) and Schaffer (2009). These works also well represent my understanding of the concept.

³⁷ Hattiangadi's contribution is particularly relevant under this aspect.

success of most of our communicative practices. For now, I will only give a simple outline of the answer.

KW's solution consists in two fundamental steps: one is to have correctness conditions for the use of a word replaced by assertability conditions, which, instead of telling us the conditions for a sentence to have a certain truth value, answer to the question "Under what conditions may this form of words be appropriately asserted?" (WRPL, p.73). One's inclination to consider their own use of the word appropriate is enough for us to say that they mean x by y , as long as -and this is the second part- the other members of their community agree with this judgment: it is, in fact, essential for one's own confidence in the way they are following a rule to be validated by the community of speakers of their language (or, more generally, of the signaling system).

One first interesting thing to notice in KW's answer is its unusual roughness: at no point is there a precise formulation or a clear thesis. The account presented is more closely related to a picture of sorts, a series of observations that are supposed to let us see that some things are there even though we cannot really frame them in words. In a way, that is the only possible methodology compatible with his "skeptical denial of our ordinary assertions"³⁸. Kripke seems persuaded by this, but I have some doubts, which I will express later on; for now, let it be sufficient to say that a meaning indeterminist answer need not be vague or sketchy just because it refuses to accept meanings as entities in the world, or as mental states.

A second important point that immediately stands out is that KW's solution is entirely *descriptive* in nature: "We will not reason [...] about the role such statements *ought* to play; rather we will find out what circumstances *actually* license such assertions and what role this license *actually* plays" (pp. 86-87). A way to understand this statement is the following: a *straight* solution to KW's problem aims at a justification for our meaning something rather than something else, that is, it looks for correctness conditions for the use of words; a *skeptical* solution, such as KW's, maintains KW's conclusion that there *really* is no meaning to look up to, but describes the *assertability* conditions under which meaning talk surfaces in actuality. This is arguably meant to show both that meaning talk has some sort of fictional nature, and that we can make sense of it without admitting to the existence of meanings. Though this is a good argumentative move, I will argue that KW's solution is not correct under a descriptive point of view either.

Finally, the infamous "social" aspect of meaning attribution. KW seems sympathetic with the idea that we can rightfully say that someone used a word in the wrong way, that is, they did not get its meaning, only in the context of a community of speakers. In assembling this argument for KW, Kripke draws fully from Wittgenstein's "private language argument": "It is not possible to obey a rule 'privately': otherwise, thinking one

³⁸ WRPL, p. 69.

was obeying a rule would be the same thing as obeying it" (Wittgenstein (1953), §202). We can get this point more vividly by considering the difference between talking to oneself³⁹ and talking to someone else. Only in the latter case we have a possible opposition to our use of words: only in a dialogue it can happen for us to *question* the way we used a word. Thus, KW points out that correctness conditions can only transpire from a social use of words, not from one's own internal picture.

Before evaluating KW's solution, it is important to assess whether we should read it as a factualist or non-factualist one, and whether the sentences in KW's account are to be considered as "truth-apt" or not⁴⁰. A factualist interpretation of KW's solution will grant that we can assign a truth value to the sentence "I mean x by y " due to a fact in the world that grounds it: for instance, if we agree that the relevant fact in question is concerned with the agreement between speakers, or with the ability of speakers to coordinate with each other, then the metasemantic sentence *corresponds* to truth. A non-factualist reading can take two main directions. One is to be a simple denial of factualism: this discards the idea that there are corresponding facts of the matter for us to determine the truth-value of the metasemantic sentence, usually associating this position with meaning non-realism. This is the account many find difficult to even consider, since it asks us to believe that meaning-talk is entirely fruitless. The alternative direction for non-factualism is arguably more promising, and it seems to have strength as a good exegetical option for KW's solution. In this case, non-factualism is associated to a deflationist⁴¹ approach to truth-value attribution: as Guardo (2020b) observes, "When Kripke says that Wittgenstein proposes a picture of language based on assertability conditions, what he means is not that truth has no place in Wittgenstein's metasemantics; all he means is that truth conditions, understood as correspondence to facts, play no role in Wittgenstein's view of language". Thus, on this reading, KW suggests it is possible to argue that " x means y by their words", even without meaning-facts to support the statement. A possible way to support this is by the quasi-realist approach we have sketched earlier.

It may be surprising to some, but nothing in what we have said so far implies the correctness of one of the three positions: the absence of meaning-I does not automatically prevent the possibility of having other facts as the candidate to determine the truth of sentences concerning language talk. Also, it remains to be seen whether either member of

³⁹ Here and, in general, when considering soliloquies or other forms of one-person linguistic interactions, I will assume, for the sake of the argument, that the person-ness is *steady* and *constant*. In reality this seems to have many counterexamples: but in such occasions (such as when I think to myself whether I am using a word properly), one can consider the multiple competing "aspects" of the self as if they were multiple interlocutors in a dialogue.

⁴⁰ Very useful for understanding this discussion are Guardo (2020b) and the usual Kusch (2006).

⁴¹ Deflationism is usually framed as the idea that for a sentence to be evaluated as "true", all it needs is to be asserted positively by the speaker. This "deflates" the power of truth-values, as they can be simply bracketed in most scenarios, which is to say that "to assert that a statement is true is just to assert the statement itself" (Stoljar & Damjanovic 2010).

the two dichotomies (factualism/non-factualism; corrispondentism/deflationism) is untenable: in such case, attributing such understanding to KW's account would result in its automatic rejection.

The Davidson incognita

Donald Davidson has frequently discussed KW's argument and his philosophy of language has been compared to the fictional philosopher's many times, with varying results. It is generally recognized that Davidson did not consider the problem in his 80s writings, then "unfairly dismissed"⁴² it in his 90s writings and, finally, gave a proper answer in his last years of activity. Most commentators, such as Kusch (2006), Sultanescu & Verheggen (2019) and Khani (2019), agree that interpreting Davidson's answer is not an easy task. The first reason is that Davidson's production is vast and complex, giving a hard time to its commentators to find an unambiguous understanding of it. The late Davidson, though, seems to agree with KW on the idea that the standard notion of meaning is fundamentally flawed and needs some conceptual engineering to adjust it, but does not accept the meaning indeterminist conclusion as presented in WRPL. In fact, he summarizes KW's own solution as "going with the crowd" (which, though crude, is not far from the truth as we have just seen), and argues that something is missing from this picture, something he calls "understanding". Such understanding is accounted for, instead, in his notion of "triangulation".

It seems to me that Davidson's triangulation argument is actually very close to the naturalistic perspective this work will suggest in its last section: if the debate with dispositionalists and primitivists has a substantial point of contention (the metaphysical existence of meanings), the debate between meaning indeterminists and defenders of Davidson's position seems to reduce itself to a mere matter of attaching different labels to equivalent facts. Essentially, Davidson and the meaning indeterminist are looking at the same facts in the world but describe them in different ways. It remains to be seen who tells the story better. In fact, as we shall see, both views naturalize meaning by reducing it to communicative strategies and attribute great importance in its workings to the influence given by the interlocutors and by the world they share with the speaker. Since this is not evident to all, I will briefly discuss Davidson's answer in the final section, attempting to assess whether my view can be seen as a close relative to it or not: my claim will be that it is, and that the conflict is only apparent.

In particular, two aspects of Davidson's position seem to be especially attractive: one is that KW's solution, as presented in WRPL, does not really have the means to answer his own challenge; the other is that what is missing from KW's account, and what a meaning indeterminist really needs to give a satisfying solution to the problem, is to explain how we *understand* each other, instead of just behaving in the "leap in the dark" sort of way

⁴² Sultanescu & Verheggen (2019), p. 8.

KW implausibly proposes in his own solution. Naturally, the two aspects are closely related: KW can't give a satisfying answer to his challenge because his communal account cannot give us a good reason to explain why communication, as we know it, works out just fine. Davidson's solution to the "aspect problem", though criticized by some (such as Khani 2017), holds the keys to a big part of the overall answer, which will find full expression in his notion of "triangulation".

A problematic aspect of the relationship between the claim I will support and the late Davidson's is that Davidson refuses to call his solution to KW's paradox a meaning indeterminist one: he argues that it is a *straight* answer, one that explains what meaning is and how it works. Discussing this idea brings up the problem of distinguishing exactly what the literature has come to call a "straight answer" as opposed to a "skeptical" (that is, meaning indeterminist) one. The most desirable outcome, in my opinion, is to count Davidson's (and our) position as a straight answer if by this we mean we should be factualists about meaning-talk⁴³; otherwise, if we mean that there is no way to rescue the ontological status of meanings, then it should count as a so-called "skeptical" answer⁴⁴. Again, a matter of terminology.

Due to this, I believe that focusing on this debate too much is counterproductive: under the metasemantic perspective, I will accommodate Davidson's refusal of meaning indeterminism by arguing that there is little difference between accepting the existence of meaning after analyzing it in new, naturalized terms (as Davidson does), and rejecting the existence of meaning as commonly understood (as meaning indeterminists do). On closer inspection, Kripke doesn't strictly deny that there could be a proper use of expressions such as "I mean x by y " or "The word 'apple' means apple": "We merely wish to deny the existence of the 'superlative fact' that philosophers misleadingly attach to such ordinary forms of words, not the propriety of the forms of words themselves"⁴⁵.

Under the philosophy of mind perspective, there seems to be no problem instead, since Davidson seems perfectly favorable to agreeing that there need be no mental objects for a mental state to occur⁴⁶.

Conclusion

We should have by now set the stage precisely enough to explain more in depth what the purpose of this study is. The idea is fairly simple: there is reason to believe -as I will try to show and the recent literature on Kripkenstein has been stressing more and more- that a

⁴³ See, for instance, Sultanescu & Verheggen (2019): "A straight solution would [...] provide truth-conditions for our declarative sentences, that is, 'conditions-in-the-world' that would make sentences true if they obtained".

⁴⁴ See, in this case, Guardo (2018): "A skeptical solution concedes that the skeptic is right about the non-existence of [meaning-I]".

⁴⁵ WRPL, p. 69.

⁴⁶ "The only *object* required for the existence of a belief is a believer" (Davidson 1997).

naturalistic and, in particular, evolutionary reading of meaning is the best candidate to help make sense of the problems Kripke has brought to the surface of philosophical debates in the 1980s. Essential to this is the distinction between a logical and a psychological account of meaning: the first frames meaning-I as an infallible lasso that catches a perfect function, that is, a function that has a correct answer for any instance of application of a linguistic signal; the second frames communication as a product of social and biological evolution, making our adherence to a rule grounded in our behavioral similarities and in the environmental stimuli that trigger them.

The one answer that fits in this framework without excessive variations is the meaning indeterminist, or skeptical, answer: there *really* is no meaning-I in the world, just animals trying to coordinate with each other in more or less complex ways, and meanings-S, if they exist, have no role in this process. Unsurprisingly, the most complex (and the most successful) of these attempts are human signaling systems, which have evolved and refined themselves into what we call “languages”. But although this is the correct solution to the riddle, there must be a reason why so many philosophers have written against it⁴⁷: my stance will be that they have focused on an aspect of this complex fact that is human communication, mistakenly taking it to be the aspect that grounds our intuitive idea of meaning something by our words, of semantics.

My main contentions will thus be:

1. That KW is right in saying that meaning-I does not exist, but, as Guardo (2019) explains and I will argue for in Section 4, this is no reason to reject the existence of communication. In particular, I find it key to accepting this idea the switch from a logical perspective on meaning to a psychological one;
2. That dispositions are effective descriptions of how *desires* work, not of how meaning-I works. In particular, I claim that dispositionalists may be so interested in this mechanic because they mix up our desire to produce a reaction in our interlocutors through our words with the meaning of such words. The confusion, in this case, is thus best explained in the field of the philosophy of mind, because they confuse one intentional state (desire) with a “fictional” or non-existent one (meaning-I);
3. That primitivist solutions are the victims of a naturalized language game, in which we *fictionally* take words to represent “stable” objects (meanings) because this in effect greatly reduces the effort needed to make others understand what we want them to understand. Such meanings may or may not be there, but the straight connection to them is not actually there. Again, moving from a logical to a psychological standpoint makes the “primitivist” perspective actually compatible with meaning indeterminism, as the primitive reference turns out to be an

⁴⁷ As Guardo (2020, p.11) suggests.

evolutionary effect on our biology instead of a logical connection between words and meanings-S.

Of course, the account I will present is not going to address every notable solution the literature on KW has put forth, mainly for brevity. That said, I do believe that at least some of them find their place in this project of “naturalization”, especially the so-called “Simplicity”⁴⁸ answer. The “Platonist” answer will also, in a way, be dealt with, if the account I am about to propose is correct.

⁴⁸ “Simplicity” argues, in a nutshell, that we mean *Chair* instead of *Chair-2030* because *Chair* is a *simpler* rule. The most evident problem here is that many rules we intuitively follow are not necessarily simpler than their “skeptical” counterparts, assuming we can (always) produce a method to determine which rule is the simplest among a set of rules.

2: The Dispositional Answer

It is now time to start focusing on the most popular among the answers to KW: dispositionalism. After all, it is no surprise that Kripke spends many more pages on this proposal than he does for any other: he probably predicted it would be the most intuitively attractive to many. One reason for the success of this theory can be found in the “standard story about past usage’s inability to determine reference” (Guardo, forth.): since our past usage of a word is not sufficient to decide which meaning-S I am following, the totality of answers I gave in the past, would give now and would give in the future if I found myself in the appropriate situation, should get the job done.

Since this is -at least by number of supporters in the literature- the most powerful meaning determinist position, I will reserve to it a befitting amount of attention. This section will thus be comprised of three main focuses: an overview of the different dispositional answers advanced in the literature, especially in recent years; an offensive aimed at their rejection as proper answers to KW; finally, an account of how dispositions can be integrated in the naturalistic account of communication by changing their role from that of an analysis of meaning to that of an analysis of *desires*.

First of all, it is now a good time to clarify the very nature of semantic dispositionalism. Coherently with KW’s concerns, the dispositionalist answer must not be understood in epistemological terms (as does, for instance, Davidson 1992). Dispositions are not, at least in this context, a device to infer, among other meanings, which one our interlocutor (or ourselves) really picked. On the contrary, dispositions are used to ground the *existence* of meaning-I and/or the nature of meaning-S: once again, it is a constitutive thesis, or at least it should be understood as such, because otherwise it becomes the answer to a different, and less fundamental, question. Usually, the philosophical analysis that is associated to dispositions explains meaning in terms of counterfactual conditionals: a speaker means *X* by *Y* if, given a certain situation, the speaker would use word *Y* to refer to *X*. The function $F(c,w)$ ⁴⁹ that (correctly) connects said counterfactual conditions with a word for speaker *a* is meaning-S: the mental state speakers find themselves in, in those cases, is meaning-I. Thus, if I expect to say “chair” when presented with a chair in 2031, what I mean by “chair” is not what rule *Chair-2030* dictates⁵⁰. This is the classic picture of semantic dispositions: any theory that wants to defend dispositionalism is theoretically allowed to make its restrictions looser, or to add new rules and conditions, if that helps make sense of it. The one component that seems indissoluble from the notion of dispositionalism itself is a *conformity* between the rule associated to the meaning-S *Chair* and the speaker’s

⁴⁹ Where *c* stands for “counterfactual” and *w* stands for “word”.

⁵⁰ I will not oppose the idea that dispositionalism can *rule out* some rules as the proper meaning of our words: assuming meanings exist, that is arguably true. The much harder problem for the dispositionalists is to argue that they are referring to rule *Chair* and *not another one*.

dispositions when using the word “Chair”: I mean *Chair* if I am disposed to conform to the rule of *Chair* and no other. Ironically, as we saw in Section 1, this gap between what the speakers are actually disposed to do and what they ought to do, if they want to follow a specific meaning-S, is also the greatest challenge to dispositionalism: it seems, then, that a good dispositional answer needs to have the solution to this problem as its top priority.

It is also useful to notice that in most cases dispositionalists implicitly have the metasemantics question in mind when discussing KW: they usually explain the reference of words by the disposition of speakers to use them in such and such ways, and they usually ignore or make little room for the question about the nature and existence of meaning-I. This approach seems to be a theoretical necessity, since the dispositional argument has a lot more strength when it accounts for the nature of meaning-S, rather than when it is applied to meaning-I considerations. This is because, as Zalabardo (1997) and Guardo (2020a) argue, the question in philosophy of mind is concerned with a mental state we supposedly have direct access to, meaning-I, whereas dispositions are inferentially known and, as such, not knowable directly and without further facts: “Justifying my present response would require forming a hypothesis about my past dispositions. But this is not how I proceed. I do not record and investigate the past physiology of my brain” (Zalabardo 1997, p. 479). Thus, meaning-S seems to be better suited as the object of analysis in this case, as it is also only knowable through inferential thought: in order to determine the rule that is being followed, in fact, one clearly needs at least some empirical evidence first, assuming, *contra* KW, that this is enough to be inferentially able to get to a plausible answer.

I think Zalabardo and Guardo are right, and that this is an important distinction to make, but with a specification about the nature of dispositions, as I understand it. Dispositions *in general* are inferentially known, because I need to make an assumption about what I would do in such and such counterfactual scenarios, on the basis of previous behavior, for instance. But some *present* dispositions are most likely *not* inferentially known: if I realize that *now* I am disposed to reply with 6 to the question “What is 3+3?”, and I do it (or refrain from doing it), it doesn’t seem that any inferential thought has been disturbed⁵¹. Since, as I will assert at the end of the section, I believe dispositions to be better suited as an analysis of desires than of meaning in the context of communication, this makes perfect sense: in the simplest cases, we don’t need inferences to know what we want *now*, but we do need to infer from our usual behavior, our usual desires in the past, or other further facts, when the present disposition is quite complex (such as a long addition) and when we consider what we *will want* in the future or what we *would want* in counterfactual

⁵¹ This is still not enough to make current dispositions a good explication of meaning-I, because *one* disposition can apply to infinitely different meaning-I mental states. Even dispositionalists would probably agree with this; if not, some of the arguments I will present against dispositionalism should be enough to answer most rebuttals on this topic.

scenarios. Though due for conceptual scrupulousness, this specification does not help in any way the dispositionalist: that is because, even though simple present dispositions are known directly, they need to hold a straight connection with the ones we only know inferentially, otherwise we may as well argue that they are not dispositions that cluster around the same meaning-S; this is but one of the instances where dispositionalists will need to show why some dispositions should be taken as part of a same set, instead of as separate entities.

With that, I leave the discussion about desires for the final part of this Section, and will take it that dispositions are, in general, not a proper approach to understanding meaning-I, if it exists.

The one-sidedness of dispositionalism exposes the theory to the danger of confusions and misunderstandings, but at the same time, by Principle of Charity, I believe we ought to sort out which question the authors are trying to answer from the start, and then assess their theory *only* in those terms. For instance, Warren (2020) is clearly trying to answer the metasemantics question: “Whether applied to rule-following, or meaning, or content, the challenge here is *metasemantic* or *metaconceptual*”⁵². Again, although this is only a part of the problem, there is nothing wrong with focusing on it: thus, when discussing his theory, I will only look at it as an answer to MS.

It seems that dispositionalism doesn’t necessarily argue one way or the other, in regard to the dependence relationship that exists between dispositions and meanings-S: it is not clear, in particular, whether dispositionalists should see meaning-S as grounds for dispositions or the other way around, even though many of them tend for the latter option, since they attempt to *reduce* meaning-S to what we are disposed to say. After all, the grounding choice turns out to be irrelevant to their claim: all the dispositionalist needs is to show that there is a Supervenience relation such that there can be no meaning-S change without dispositional change, i.e., my dispositions about the use of “chair” can change iff I *mean* something else by that word. Thus, I understand the dispositional account as *relational* (since it claims the existence of a one-to-one relation between dispositions and meaning-S) but not automatically *hierarchical* (since there seems to be no important difference whether the more fundamental fact be meaning-S or our dispositions).

We should also discuss the nature of KW’s objections to dispositionalism, because their relation and similarities are not entirely clear. Guardo (2018) argues that *Error* and *Finitude* are two sides of the same coin, which he calls “unreliability problem”⁵³: essentially, both *Error* and *Finitude* are ways to show that we cannot *rely* on what we are disposed to do in certain circumstances in order to determine what we mean by a word. Doing that would

⁵² Warren (2020), p. 258.

⁵³ Guardo (2018), p. 23. My translation from Italian (henceforth, *Unreliability*).

amount to accepting that the correct answer to the sum of two numbers, so big we don't have enough surfaces to read them on, is a shrug. But, of course, this is completely unintuitive: we know our inability to answer is given by environmental, biological and behavioral causes, not by the rule of addition itself, which works for any two addenda, however huge they might be. Take the function $F(c,w)$: a meaning-S attaches one and only one word⁵⁴ (or its negation) to each conditional. For instance, when presented with a chair, rule *Chair* will attach the word "chair"; when presented with a table, rule *Chair* will attach "not-chair". *Finitude* and *Error* seem to show that our dispositions frequently drift from this rule: specifically, *Finitude* shows that we don't really comply to the whole function (because sometimes we don't know what to say, for example), while *Error* shows that even when we do comply, we don't always comply to *that* function (because sometimes we might call "chair" a table by mistake). The big problem posed by this picture is that the dispositionalist needs to find a way to explain why a mistake is a mistake, instead of the correct rule-following of a *different* meaning-S.

Now, I believe there is a strong connection between the *Normativity* objection I presented in Section 1 and the point we have been making here. If we look at the reason why *Finitude* and *Error* are powerful objections to dispositionalism, it seems that we find our answer in the discrepancy between what we do and what we ought to do: when being asked "What is 2+3?", we *ought* to reply "5" if we want to follow the rule of addition, but we *may* actually reply "6" or not reply at all; hence, it seems that we could understand *Finitude* and *Error* as two instances, two examples, of the general objection to dispositionalism that *Normativity* represents. To put it differently, *Finitude* and *Error* point out a problem with the *extensionality* of dispositions: while the addition of 3+2 refers to, and only to, the number 5, our dispositional answer to the arithmetic question can refer to anything at all, as long as we can have the disposition to refer to it. Since the function representing my dispositions is not the same as the one encapsulated by the meanings-S rule, meaning-S cannot be explained in terms of dispositions. *Normativity* explains why that is the case: because dispositions "happen" to be the way they are and cannot be, in and of themselves, "wrong", while meanings-S are norms to follow. Dispositions cannot be wrong by themselves, precisely because they need to be *compared* to some set of answers (meaning-S) in order to determine whether they are appropriate or not.

To sum up, then, this is the basic understanding of the Normativity problem that dispositionalists need to deal with. The one attack KW needs to launch on dispositionalism, and that seems more than enough to destroy it, is the following: dispositions are *descriptions* of what we do in the current context or would do in some other context; meanings-I, in the strong sense meaning-determinists defend, are attempts at *rule-following*, thus their effectiveness depends on a *prescription* (meaning-S) we can always *fail to comply with*. Ergo, for KW, dispositionalism is not just an incorrect

⁵⁴ Consider synonyms as "the same word".

philosophical explication of meaning-S, but it is also completely beside the point of meaning-I, thus being wholly useless as an attempt to solve his problem.

This said, it is important to realize that the literature has given multiple interpretations and novel readings of the relationship between the three problems. Though I believe the one I just presented to be the most faithful portrayal of KW's position, then, one ought to tread carefully when finding out what a specific author believes to be the "Normativity" problem before addressing their arguments. For instance, some philosophers⁵⁵ try to answer KW's problem in dispositional terms without addressing Normativity at all, because they see it as an extension of *Error*, while leaving *Finitude* as a problem that stands on its own.

The reason for such different perspectives on the correct reading on this matter seems to be that there is more than one plausible understanding of the *Normativity* claim: one can reasonably see *Finitude* as a separate problem if they don't interpret the lack of answer as an answer itself, but rather as a suspension of judgment, for instance. This is not to say that all interpretations of the problem are "born equal", but rather that assessing which is the correct one may be harder than it seems at first glance.

Because of this, it seems useful to take a closer look to the three main understandings of Normativity that the literature on Kripkenstein has produced over the years; this will also grant us a precise terminology for arguments that will surface later on.

Normativities

Normativity-1: The "naïve reading"⁵⁶. This is, in my view, the most straightforward formulation and the closest one to what Kripke actually wants to say in WRPL, due to statements such as the following: "As a candidate for a fact that determines what I mean, [the dispositional account] fails to satisfy the basic condition: that it should *tell* me what to do in each new instance. Ultimately, almost all objections to the dispositional account boil down to this one" (WRPL, p. 24).

There is a fundamental gap between a dispositional state and meaning-I: the first one has a *descriptive* nature, as it describes what we would do in certain situations, whereas the second one has a *prescriptive* nature, in that it tells us what we *ought to* do in order to comply with the meaning-S we want to convey. Because of this difference in nature and function, dispositions are not an appropriate instrument to analyze meaning. In Paul Boghossian's (1989) famous formulation, *Normativity-1* is "simply a new name for the familiar fact that [...] meaningful expressions possess conditions of correct use" (p. 513). Notice that in this reading there is no obligation to use a certain word other than for the

⁵⁵ Such as Båve (2019).

⁵⁶ This name is my own: it implies that it is the most "immediate" understanding one usually gets from reading WRPL.

intention to use it in a certain way: if you mean *Chair* and you want to use the word "Chair" in a way that fits what you mean, then, you have to use the word "Chair" in such and such a way; if not, you can also deceive the interlocutor or use a new word altogether, unlike in the *Normativity-3* understanding.

Normativity-2: The "Privilege argument"⁵⁷. After outlining the naïve reading, Boghossian addresses an interesting objection to it: "What if, instead of identifying what I mean by 'horse' with the *entire* range of my dispositions in respect of 'horse', it identified it only with certain select dispositions [?]"⁵⁸. This, as we will see, is not a good answer to the problem for a few reasons, but it yielded a new facet of the Normativity considerations, and a new powerful objection to dispositionalism: we have no reason to prefer some dispositions than others, because that would seem to imply we are privileging some dispositions in a completely arbitrary way, i.e., there is no *norm* capable of having us decide which dispositions should be preferred.

It seems we are in the presence of a completely different understanding of where the "Normativity problem" lies: in *Normativity-1* the "Normativity problem" refers to the fact that the nature of dispositions is not normative; here, on the other hand, the "Normativity problem" refers to the fact that we have no norm to determine which dispositions are to be chosen over others. At the same time, *Privilege* can also be understood as a "concrete" application of *Normativity-1*: it challenges dispositionalism, especially in its idealized version, to give us a reason why some dispositions are the "right" ones, the ones we should follow.

Normativity-3: The "Ought argument". Some philosophers read *Normativity* in a stronger way, that touches on the actual actions of the speakers. This form of normativity argues that one ought to use words in the correct way, that is, in the way other speakers will understand. This is analogous to the meta-obligation to comply to moral rules: one is not merely obliged to follow a moral rule *if* they want to be a part of the "morality game"; one has the obligation to participate in the morality game in the first place. Similarly, one does not simply have to follow the *Chair* rule *if* they want to talk about chairs; they are *obliged* to do so because it is obligatory to be a part of the language game of one's community. I claim that this reading is completely beside the point: KW never discusses this sort of normativity, which seems much more concerned with social ethics about communication and not at all with the metaphysics of meaning-I or of meaning-S.

As it is apparent, the different readings are *very* different, and there are subtler shades to each of them. Depending on the preferred understanding of Normativity, then, there will be different implications in the arguments for or against KW. As such, an author that defends the *Normativity-2* reading, for instance, would be justified in not accepting my

⁵⁷ As labelled by Guardo (forth.).

⁵⁸ Boghossian (1989), p. 532.

view of the relationship between the three basic objections to dispositionalism. That said, to prevent further confusions, I will normally refer to *Normativity-1* with “Normativity” and call *Privilege* and *Duty* the other two readings when useful for the discussion: that is because Normativity-1 is arguably the version KW is closest to, but other understandings, especially *Privilege*, are still useful in the debate as independent claims.

SDA, again

We have introduced the Simple Dispositional Answer in the previous section, but only in terms of WRPL. Naturally, since then, some new attempts to validate it have been tried out, though even dispositionalists themselves seem to have now lost faith in this approach. When confronted with the notion of conformity to a rule, SDA simply states that the dispositions we have determine the meaning of a word: thus, if I am disposed to call “chair” chairs, I am following the *Chair* rule. SDA is especially hard to defend, because, due to its extreme simplicity, it seems to fall prey to KW’s three powerful arguments, which we have subsumed under the labels of *Unreliability* and *Normativity*.

A plausible way to try to save SDA would then be to object to KW’s arguments used to reject it (*Error*, *Finitude*, *Normativity*). Blackburn (1984b) is arguably the strongest defense of SDA, so let us look at its main points of attack. Blackburn points out, against *Finitude*, that “It is not obvious that dispositions in themselves are either finite or infinite. The brittleness of a glass is a respectable dispositional property. But there is an infinite number of places and times and strikings and surfaces on which it could be displayed”. Blackburn is not wrong when he says we may very well have infinite dispositions. Recall that dispositions do not need to be actualized in order to “be there”: they would be actualized if the speaker that holds them found himself in such and such situation, which is exemplified by the counterfactual conditional. So, it actually seems much more likely that dispositions *are* infinite than that they aren’t: say you are asked your name by a person; say that person has a specific number of hairs on their head; say their hair have a specific medium length; say their eyes have a specific shade of color... These countless features could differ in infinite ways in counterfactual conditionals, but they presumably would not change your disposition to state your name when asked⁵⁹.

As Guardo (2018) points out, though, it only takes a rephrasing of Kripke’s point to give it new strength. Thus, it must be admitted that Kripke’s first formulation⁶⁰ of *Finitude* does not really work, because it assumes the finiteness of dispositions, but his later reformulation does, because it doesn’t assume any of that: the problem turns out to be that “Dispositions cover only a finite segment of the total function [they refer to according to

⁵⁹ One could argue that the reason why these infinite variables don’t change the disposition is that these variables are irrelevant: that is beside the point here, because we are merely assessing whether dispositions can be infinite or not.

⁶⁰ The one I quoted in the first Section.

the dispositionalist]”⁶¹. Basically, even if we accept the (after all, not unintuitive) idea that our dispositions are infinite, the problem is that they are not all adherent to the rule we intuitively think we are following. For instance, say you are blindfolded and asked what piece of furniture is in front of you. If it is a chair, a correct use of the meaning-S *Chair* would be to say: “It is a chair”. Unfortunately, being blindfolded, this instance of application of the rule (and infinitely many others!) just isn’t contemplated in your dispositions, as it is more likely to consist in a shrug or a confused expression. Thus, it is clear that, even though the function “Given condition x , I am disposed to say y ” is actually infinite, it is a different function from “Given condition x , I ought to say y to conform to meaning-S”, and arguably a smaller one anyway⁶². Kripke focuses on the *time* needed to determine the correct use of a word⁶³, but the examples are analogous: in general, a biological or environmental condition prevents the speaker from being able to give an answer that corresponds to the meaning-S they supposedly mean to express.

Blackburn’s other line of defense consists in claiming that people are disposed to correct their mistakes when they are made aware of them, so errors can be seen as a physiological aspect of following a rule. This remark is somewhat similar to those around which ideal-condition versions of the dispositional analysis revolve⁶⁴, but not exactly: we remain in the realm of the *actual*, and that is exactly why this objection is a lot weaker than those ones. In the actual world, many people (maybe *everybody*) make mistakes in some fields of knowledge and are not cognitively able or never find themselves in the condition to recognize their mistake as such. Blackburn’s argument doesn’t account for mentally retarded people, for instance, who, to some extent, are simply unable to understand mistakes and to correct them. Does this amount to saying that they don’t mean *Chair* by “Chair”, or addition if they say that “ $3+3=7$ ”, while more cognitively capable people do? Intuitively, it seems not.

To conclude, we may want to look at a possible state of affairs where all my dispositions actually mirror exactly the *Chair* function. Though extremely unlikely – and, because of that, almost certainly not actualized or possible in the real world – it may indeed be the case that a pretty overconfident individual is always disposed to answer in the exact “Chair – Not-Chair” dichotomy expressed by *Chair* (no *Finitude*) and that, by pure chance, they always get it right (no *Error*). The capital question here is: In a state of affairs where *everyone*⁶⁵ has this kind of dispositions, would semantic dispositionalism be correct? I don’t think so, and it is time to explain why by returning to the Idealized Dispositional Answer.

⁶¹ WRPL, p. 32.

⁶² If seen as transfinite infinities. That is because while meaning-S has a binary answer (“Chair” or “Not chair”) for any possible situation, it is not clear that we are always disposed to give one of said outputs in any circumstance.

⁶³ WRPL, pp. 26-27.

⁶⁴ And, in a way, to what Warren (2020) will suggest later on.

⁶⁵ Or, at least, the majority of speakers.

IDA, again

The Idealized Dispositional Answer is clearly stronger than its predecessor. This strength can be traced back to its more complex counterfactual conditionals. At least *some* of the objections SDA falls prey to, don't seem to work as well here. As I was hinting at, IDA basically silences, by premise, all the "environmental noise" that is culpable for our lack of adherence to the meaning-S function we want to convey by our words. This is somewhat intuitively attractive: returning to the example from Section 1, if I call Mark by his twin's name, "Paul", upon realization I will be inclined to say "Sorry, I *meant* 'Mark'!".

Importantly, the subtext of this expression seems to be: "In a situation where I saw more clearly/where I was not intoxicated/where I noticed your birthmark/etc., I would have correctly called you by your name". This is what the literature usually calls "*ceteris paribus* clause": what I really mean by my words is what I would say in ideal conditions, i.e., where I don't make mistakes.

Precisely because this idea is so intuitively strong, it is important to look at it with more precision, as Kripke brilliantly predicted by giving this suggestion a comparatively long refutation in WRPL⁶⁶. Kripke gives two arguments against IDA: the first is concerned with the problematic nature of the idealized scenarios the dispositionalist may choose, the second one with the circularity of idealizations. The two arguments are not entirely apart in Kripke's mind: rather, we might say that the second one is a rebuttal to whoever thinks there may be a solution to the first. As I claimed in the first Section, I believe the first argument to be wrong on its own terms, while the second one is decisive to reject IDA. It is now time to see why.

As anticipated, Kripke is not convinced by the idea that we can predict what we would do in "alternate reality" conditions: having a huge brain or being able to live 500 years in order to be capable of answering correctly to an enormous addition may have unexpected effects on our dispositions, so the mental experiment, in his view, doesn't tell us anything in favor of the dispositional account. It seems to me that there are three problems here, with the third one being the nail in the coffin: (1) Kripke chooses *ad hoc* sci-fi level examples; (2) The idealization is supposed to be able to rule out any possible unpredictable outcome *a priori*; (3) He gives a skeptical account of our predictable behavior. Let us see them in turn.

(1): Guardo (2018; forth.)⁶⁷ points out that there are much more plausible alternative states of affairs that don't lead to such problematic situations: for instance, I may not be disposed to give a proper answer to an addition because I can't remember the numbers in the addenda. A mere piece of paper and a pencil are sure to suddenly make the job a piece of cake, though. Analogously, my being unable to tell whether I am in front of a chair or a

⁶⁶ Pp. 27-31.

⁶⁷ P. 47 (2018); p. 14 (forth.).

table because I am blindfolded would be resolved by someone pulling the blindfold off my eyes. No great leap of imagination is needed here, so the “mysteriousness” of the ideal conditions seems to disappear for most cases. Mind you, there is nothing unwarranted *per se* in Kripke’s examples: they are simply too easy an escape route from the dispositionalist attacks. By the Principle of Charity, we should give the dispositionalist as much room to defend her position as possible, and the alternative ideal conditions are of equal conceptual might, without the problematic further questions that arise from Kripke’s examples.

(2): Let us assume what Kripke says is correct: we have no way to know what would happen if more grey matter was inserted in our brains or what we would be disposed to do with a much longer life. This seems beside the point of idealization, because further idealizations could rule out such unpredictable and undesired outcomes: if an exaggerated amount of grey matter would lead us to insanity, idealize so that this is prevented; if living millions of years changes one’s perspective on how to approach additions, idealize so that one has enough time to calculate the addition without excessive existential turmoil.

(3): Even in those cases presented by Kripke, it is not entirely clear that we have no way to predict what would happen, because I suspect this might be a case of overzealous skepticism. Surely, a team of neuroscientists would be able to predict (to the best of our knowledge) what would happen if our brain was bigger or, more precisely, *better developed*. It is a simple matter of fact that our brain makes us more capable than monkeys in intellectual activities, so a creature as better developed than us as we are compared to monkeys would *presumably* make less mathematical mistakes. A similar remark can be made for a longer life: it just seems unwarranted to claim complete ignorance of what would happen in these (for now) merely imaginary states of affairs⁶⁸. Fodor (1990) makes a related, though non-identical, remark: he points out that scientific reasoning often times employs idealizations of similar kinds, and this is not only useful, but essential to understanding the actual, imperfect world we live in⁶⁹. Kusch (2006)⁷⁰ is not happy with Fodor’s account though, and I agree that, in those terms, it may not be a proper objection.

In any case, having rejected Kripke’s first line of argument is not sufficient to think that IDA is to be accepted. Far from it, I believe the other, more general objection Kripke puts forth to be decisive and indisputable as a rebuttal to this theory. In order to show its full force, we need to employ one of the contemporary readings of *Normativity* we have introduced earlier in this Section: the “Privilege Argument”.

⁶⁸ Here we are clearly talking about inferential knowledge, which is of a different nature from meaning-I, which is (supposedly) known directly. I believe my argument holds because, as pointed out earlier, dispositionalism is usually an analysis of meaning-S, which is most likely known inferentially.

⁶⁹ Fodor (1990), pp. 94-95.

⁷⁰ Pp. 101-105.

Assume we find a way to shut down the problems given by *Error* and *Finitude*, as IDA does, thus linking a certain set of our dispositions to the meaning-*S Chair*. Even then, what makes it the case that we should decide that *that* set of dispositions is the one we should focus on, and that any other collection of dispositions is not? What makes this decision a justified one in semantic terms? Unfortunately, it seems the only answer the dispositionalist can suggest us is that this is what we really mean by our words. But how is this helpful in any way? Surely this cannot satisfy us, since the entire point of IDA was to give us a fact that would *explain* why we mean something by our words. Instead, the idealization implies that we already need to know what we mean if we want to give proper idealizations to our actual behavior and thus have the desired result: since we mean addition in the first place, then we remove the situations where we reply incorrectly to the question “What is 3+2?”; since we decided we mean *Chair* by “Chair”, we idealize so that we remove the deviant scenarios where we used the word “Chair” to refer to tables. Let me schematize to make the point as clear as possible:

The IDA project is:

- Ground our meaning *S* by word *w* in terms of idealized dispositions;
- An idealized disposition, by definition, removes all “noise” that produces, in actual circumstances, mistakes or lack of answer to a semantic stimulus;
- A mistake or lack of answer is defined in terms of a correct rule we should follow;
- Thus, IDA presupposes the “real” rule we follow, in order to ground the fact that we follow it.

In conclusion, IDA is question-begging and, thus, it seems doomed to refutation, once again, due to its very nature.

Ginsborg’s IDA

We should give one last chance to IDA through its latest prominent proponent, Hannah Ginsborg (2011a; 2011b). Her contribution to the dispositionalist strategy is to suggest a quite unusual condition for the disposition to be the right one. Such added condition we need, to have IDA work, is that the speaker must be disposed to find his own disposition *appropriate*: “What the normative proviso does is to build into your disposition the feature that every response you are disposed to give involves a claim to its own appropriateness to the context in which you make it. Your disposition is thus not just to say '125' in answer to '68 plus 57,' '126' in answer to '68 plus 58' and so on; it is also, in each case, to take what you are saying to be the appropriate response to the query. You are disposed, not only to respond with a number which is in fact the sum, but to consider that particular response to be appropriate [to the meaning-*S* you are referring to]”⁷¹.

⁷¹ Ginsborg (2011a), pp. 16-17.

The main interesting aspect of Ginsborg's account is, to my eyes, the observation that we don't always make a judgment on the way we use words: on the contrary, we frequently "speak without thinking". Far from there being anything wrong with it, we *need* to do that in order not to make every conversation an intellectual challenge to our cognitive abilities; yet, this also makes for dispositions to use words in ways that we ourselves may not agree with upon reflection. Thus, Ginsborg suggests that we consider our real dispositions only those dispositions that we would judge positively after an evaluation of appropriateness: if I call a table "chair", as soon as I notice it is not a chair, I will judge my previous disposition as incorrect and, thus, it ought not to be taken as part of the set of dispositions that make up my meaning *Chair* by "Chair". This is somewhat similar to a situation where authors write down a set of sentences and then, after re-reading them and correcting some of them, they publish their work: of course, we would not say their work is mistaken because there was a mistake in the first draft⁷².

Even so, I am convinced that this suggestion misunderstands the problem and still fails to distinguish the descriptive nature of dispositions from the prescriptive nature of meaning. Essentially, what doesn't work here is that being disposed to consider one's answer appropriate does not automatically make the answer appropriate: if an author corrects some sentences in a book, that does not *automatically* tell us that what he deems to be a correction is actually more correct than the previous draft was. Actually, it seems normal for people who make mistakes to take their mistake as the correct thing to do: otherwise, they would not be making the mistake⁷³.

Finitude also seems to give problems to this account. We cannot really say that we are *always* disposed to think our answer appropriate when we give one, so sometimes we say things that we don't mean: sometimes, for instance, we are disposed to answer in a certain way because we are pressured to, not because we are certain about its correctness. Is Ginsborg happy to say that in those instances we don't *mean* what we say? See this with an example: John says that " $3+2=6$ ". Does he not mean addition by his words? It seems much more plausible (in the meaning determinist perspective) to argue that we mean addition if we answer 125 to the question "What is the sum of $56+68$?"⁷⁴, but also that we may be wrong about the answer. How to adjust this intuition to the dispositional account remains unknown.

Say, on the other hand, that Ginsborg means that this added disposition must be *itself* idealized, (i.e., we would be disposed to deem the answer appropriate given that we know

⁷² A similar remark can also be found in Blackburn (1984) by the name of "extended dispositions", and in Warren (2020) by the name of "stable dispositions".

⁷³ Unless they make the mistake on purpose, but, then again, we could question whether that counts as a "proper" mistake.

⁷⁴ Indeed, the answer is arithmetically wrong here: but if one didn't notice, wouldn't they call this an addition still?

a priori the correct answer), then she falls prey to the “question-begging” problem just like her predecessors: the idealization cannot take for granted that the speaker means *x* rather than *y*, because it is the whole point of the dispositional account to explain why she means *x* rather than *y*. Finally, a word on Ginsborg’s somewhat mysterious notion of “primitive normativity”. She argues that this concept is vital for her theory, in that it allows us “to take [the] use as primitively appropriate to the context”. This sounds a lot like a meaning indeterminist claim, though: actually, the *psychological* feeling of appropriateness of one behavior rather than another when presented with a word is exactly what meaning indeterminist theories can claim to ground the practice of communication without resorting to meaning-I⁷⁵.

Elsewhere⁷⁶, Ginsborg explains that this primitive normativity needn’t distinguish “at least in the first instance” between correct and incorrect uses of a word, but rather between states of affairs where the question about correctness arises and states of affairs where it doesn’t. This does not seem to help as a dispositionalist theory, either. As Miller (2019) argues, it is not clear how this prevents the need for a pre-determined correct use of the word (i.e., of a meaning). Ginsborg is telling us we are “primitively” able to distinguish between two sets of responses to the use of a word:

- (1) Responses where it is appropriate to question correctness;
- (2) Responses where it is not appropriate to question correctness.

But then, she either falls in the usual regress of IDA responses (because she needs to presuppose a rule that makes it possible to decide when the correctness of some responses can be determined, whether or not they turn out to be correct or incorrect) or, again, she conflates in our meaning indeterminist answer, if this primitive “feeling” has a psychological, rather than logical, connotation.

It seems that we are in the presence of one of many replies to KW that miss the mark because they essentially agree with the meaning indeterminist claim but want to safeguard meaning in some way for the reason that they assume it to be essential to proper communication⁷⁷. Ultimately, then, Ginsborg’s proposal seems to fail where other IDA answers fail and her notion of “primitive normativity” can be seen as equivalent to a meaning indeterminist answer or as incoherent if meaning determinist. Essentially, her attempt to find a “middle ground” between dispositionalism and primitivism makes the positions incompatible with each other, thus making polarized positions more plausible than hers. Alternatively, if we are happy to read the “primitive normativity” argument as

⁷⁵ As Guardo (2019) does and I will argue for in Section 4.

⁷⁶ Ginsborg (2011b), p. 169.

⁷⁷ Kusch (2006) makes a similar remark when he argues that many self-proclaimed “primitivist” answers really are “in fact closer to meaning [indeterminism] than they are to meaning determinism” (p. 235). I think that Ginsborg’s attempt follows this tendency.

a psychological one, then we are simply turning our head to the form of meaning indeterminism I will support.

With this other option out of the way, I think we can safely assume that IDA is not the correct strategy to defend dispositionalism, if there even is one.

State of the art dispositionalism and the algorithmic approach

It seems we are, after all, justified in concluding that SDA and IDA do not work out as a satisfying answer to KW's problem. By itself, though, this still does not automatically mean that dispositionalism is doomed to fail yet: there may be better, although arguably more subtle, pictures of our "actual or [...] possible linguistic behavior"⁷⁸ that enables the dispositionalist to, at least, answer the metasemantics question, since we have pointed out how the dispositionalist strategy really seems to be out of the question as a good account for the problem in the philosophy of mind. I will now discuss the two most recent dispositional attempts: both will result misdirected, but also useful to show how dispositionalism can be implemented in our theory under a different aspect. Probably the most interesting dispositional approach in recent years is found in the, already mentioned, Warren (2020).

Warren's strategy starts from an algorithmic approach to dispositions, where simpler dispositions are molded together to form a more complex one. In Kripke's terms, the idea of an algorithmic approach to the problem is that "I learned – and internalized instructions for – a *rule* which determines how addition is to be continued". This can easily be translated in dispositional lingo: I am disposed to answer with the sum of two addenda instead of anything else, because I am disposed to follow a *procedure* that leads me there; a procedure that can be applied to any example. The procedure Kripke suggests is that of counting the number of marbles in one heap (the first addendum), then the number of marbles in another heap (the second addendum), and finally count the marbles from both heaps together (the sum). Surely, the defender of this idea would say, this procedure works out for *any* addenda whatsoever, so it cannot be confused with any other rule. Unfortunately, Kripke *prima facie*⁷⁹ convincingly shows that this is not true: "the sceptic can question my present interpretation of my past usage of 'count' as he did with 'plus'". Basically, what Kripke is suggesting here is that each and every word that comprises the "rule set" for additions can fall victim to the same problem as addition did: we still have no reason, no fact, to determine that "count" is not *Count-2030*, such that, from 2030 onwards, one is supposed to count in couplets; "marbles" are not *Marbles-2030*, such that, from 2030 onwards, a marble is made up of two pre-2030 marbles; and so on. This argument may sound more intuitive when taking the *Chair* example, because the

⁷⁸ Warren (2020), p. 258.

⁷⁹ As we are about to see, Warren shows that the dispositional interpretation of the algorithmic approach can work around this problem.

dispositions we have to determine what a chair is, are very similar to the dispositions we have to determine what is part of a chair: if my disposition to call a chair “Chair” is grounded in my disposition to call its legs “legs”, its seatback “seatback” and so on⁸⁰, it is obvious that each of these words can be interpreted as Goodman cases themselves.

Warren doesn’t seem as pessimistic about this strategy, though. In particular, he claims that an interesting way to salvage a dispositional account based on the algorithmic idea would be to find *external* rules instead of following *internal* ones, as our marble-counter did. The first step to work in this direction is to make a clear distinction between simple and complex dispositions: simple dispositions are such that there are no intermediate actions between their arousal and their satisfaction; complex dispositions are such that there are intermediate actions the speaker is disposed to take before satisfying them, namely simple (or simpler) dispositions⁸¹. Now, simple dispositions do not fall prey to the interpretative attack because there is no rule to interpret: the disposition is satisfied without recurring to *explicitly* semantic middle steps. After all, we don’t need to understand what the word “count” means in order to start counting. The elegant move Warren makes to defend complex dispositions, then, is to argue that they can be analyzed in terms of more or less long procedures made up of simple dispositions⁸²: for each step, there are no internal rules to interpret, ergo complex dispositions are safe as well. Now that simple and complex dispositions are exploitable, they must be shown to be consistently used over and over in the same way, which is what he calls a “composite disposition”. With that, it really seems we are able to call this the fact that determines we are following a proper rule.

Warren’s work is moreover significant because he makes subtle points about anti-dispositionalist accounts, showing new flaws that lay undetected until now. A powerful example of this is his rebuttal to a *reductio ad absurdum* argument that, *prima facie*, looks promising and equivalent to the classic WRPL considerations. Though this is probably not the correct hermeneutical reading of KW’s objection, it is still worth looking at, in order not to fall for it in future anti-dispositional arguments. Firstly, Warren shows that it is not quite obvious that the law of transitivity should apply to *explanations*, that is, that “If *a* explains *b*, and *b* explains *c*, then *a* explains *c*”. In particular, he argues⁸³ that if we don’t accept such rule, it is not warranted to jump from:

⁸⁰ Which is itself a very controversial hypothesis under a phenomenological perspective. Analogously, we could plausibly suspect the idea that we add by actually following any kind of algorithmic procedure in normal conditions, rather than by “direct intuition”.

⁸¹ “Complex dispositions are exercised by way of exercising various simple dispositions” (p.264). Complex dispositions are not to be confused with “Composite” dispositions, which Warren defines as dispositions to repeat a specific simple or complex disposition (such as calculating *in the same way as before* when asked to add two addenda).

⁸² This idea is not entirely new: see Blackburn (1984b) and Hattiangadi (2007), for instance.

⁸³ I reformulate his argument [Warren (2020), pp. 278-279] with the Goodman example.

- (1) The justification to call a chair “chair” in 2031 is explained by the meaning-S the speaker refers to (*Chair*).

And the dispositionalist principle:

- (2) The meaning-S the speaker refers to is explained by the speaker’s disposition to use the word in such way.

To the, apparently absurd, conclusion:

- (3) The justification to call a chair “chair” in 2031 is explained by the speaker’s disposition to use the word in such way.

This is a line of reasoning one could be tempted to use to object to dispositionalism: given the intuitive (1) and (2), which dispositionalists agree with, we come by *reductio* to the unacceptable idea that all it takes to be right in using an expression is for the speaker to use it⁸⁴. It is also interesting to notice that Warren frames dispositions in *explanatory* terms instead of reductionist terms, as they are usually employed. That is relevant because most dispositionalists *reduce* meaning-S to dispositions about words: here dispositions give a reason for meaning-S to be in such and such way. This seems to give room for a weaker reading of dispositionalism, that doesn’t necessarily give a grounding hierarchy to either dispositions or meaning-S.

Assume, then, that we accept the objection that transitivity of explanation should not be taken for granted here, since this is plausible at least in some other contexts⁸⁵: this is still not necessarily a problem for anti-dispositionalism. Instead of finding a rigidly logical *reductio* strategy, in fact, the greatest appeal of meaning indeterminism is of a holistic nature: so much more can be explained in naturalistic terms, because we don’t deal with mysterious entities such as meaning-I and meaning-S anymore, while still being able to safeguard communication and language. Also, premise (1), which Warren takes as “an obvious truth, central to our notion of meaning” is definitely not as obvious for KW: what the fictional philosopher would say is, rather, that the justification to call a chair “chair” in 2031 is found in the appropriateness of such utterance, as judged by his linguistic peers. This is to say that KW breaks the spell of meaning determinism at a much deeper level than what premises (1) and (2) may show: even if the logical *reductio* -one, not particularly prominent, objection to dispositionalism- is flawed, then, this doesn’t reduce the overall strength of meaning indeterminism.

⁸⁴ Which is unacceptable for the familiar Wittgensteinian reason that “Whatever is going to seem right to me [is not automatically] right”.

⁸⁵ In particular, it seems reasonable to argue that there is an important difference between different levels of *directness of explanation*: my parents’ existence is a cause of my existence, my existence is a cause of this essay being written, but my parents’ existence is not a cause of this essay being written *in the same way* as my existence is.

A central claim in Warren's dispositional solution is that we can be *fairly* certain someone means some meaning-S by their words because, *in the vast majority* of situations, they will be disposed to use meaning-S correctly. This, he argues, is already enough to eliminate errors in "normal situations". After all, we frequently accept facts in the world by arguing that they are the most likely to happen: medicines are said to be good for us in general, even though they can have negative side-effects in a minority of cases. When Warren argues against the logical *reductio* we have seen, he uses this notion profusely. For instance, he introduces a rule called "SM-Normativity", such that "If utterance α is semantically correct, for speaker S in context c, then S has semantic justification for uttering α in c". Now, this is a principle that "will be accepted by all who accept semantic normativity, whatever they think about the nature of meaning or rule-following", precisely because it is not yet specified what "semantic correctness" amounts to.

Warren then claims that if SM-Normativity is true, then the *apparently* absurd (3) will actually be true as well, because it is plausible that the frequently apt use of a rule makes the user a follower of the rule: the speaker's disposition to usually follow the rule plausibly gives his expressions semantic correctness and, thus, justification for his utterances. We may want to spell out the epistemological implications of this *probabilistic* approach to meaning attribution: if it is statistically likely enough for a speaker to use a word according to a meaning-S rule, then it is rational to believe that they are following such rule and, thus, that they mean that meaning-S by their words. Warren presents two principles that are supposed to enable us to determine one's meaning-I: Normalcy and Stability. As he notices, the notion of "normal" conditions can be very problematic for the dispositionalist, because it seems to fall for the same circularity objection that makes IDA unviable. Thus, Warren gives us non-semantic conditions to determine such state: "The normal situations [...] are those in which neither external, nor internal factors are interfering with [one's] general cognitive functioning". This makes it possible to screen off any disturbance to the correct application of meaning-I that is not directly caused by one's own cognitive abilities. But this is still not enough: we have seen, and it is quite obvious, that one can make *frequent* mistakes even in such "normal" conditions. This is where Stability comes to the rescue: "This final screen involves factoring in what happens when [the speaker] repeats the process, in effect checking and rechecking his work and retracting previously given answers when they fail the check".

The two filters to our basic understanding of dispositions ought to give us a proper selection of all and only those dispositions that must be taken into account as the facts determining what we mean by our words. After such finessing, in fact, the appropriate uses of the rule we comply with will be overwhelmingly more frequent than the inappropriate ones. Even so, I think that the idea misses a fundamental point.

The reader will remember that we have introduced an interpretation of the Normativity argument that points out how we don't seem to have a good reason to choose *some* of our

dispositions over others: the *Privilege* argument. Unfortunately, it seems that the *statistical* perspective should have no power here, precisely because, even in meaning determinist terms, there is no telling whether what Warren considers a rare “mistake” is actually a mistake or not: it may well be a correct answer from a different rule.

Let us see this by employing our usual example:

I intuitively expect to be following the *Chair* rule. One day, at a postmodern design museum, I mistake a table for a chair and thus I use the word “chair” to refer to the table. Warren finds this acceptable, because I never make this kind of mistakes in normal situations: the unusual shape of the piece of furniture is responsible for the unusual response I am disposed to give, so my meaning-I *Chair* is safe. But, if we apply *Privilege*, this does not seem true. There is still no fact of the matter for me to know that I was following *Chair* and not *Chair-museum*, a rule such that:

In general, follow rule *Chair*;

In museums⁸⁶, swap rule *Chair* with rule *Table*.

Once again, it is important to notice that the -at least apparent- absurdity of this proposal is not meant to argue that this *is* the rule the speaker was actually following: it is meant to show that talk of meaning-I is simply nonsensical in general, that there is no such thing as strict rule-following in language, precisely because we can make any signaling decision accord with as many different rules as we can come up with, as long as they are compatible with the overt behavior of the speaker. It may be useful to show what I mean by making a comparison *by difference* with the skepticism about our perception of the real world: I believe this comparison to be essential in showing that the doubt about meanings has warrant, whereas the one about the outside world doesn't, thus making only the second a proper form of “skepticism”⁸⁷.

An argument some philosophers use to defend the reliability of our perception of the world consists in claiming that the instances where our perceptions deceive us are few and far between: *usually*, they represent reality exactly the way it is⁸⁸. One can easily notice the striking similarity the skeptic about our perceptions has with KW's, and the reason why KW is so often associated with this philosophical current: we have an obvious intuition, and the skeptic disrupts its solidity with the devastating power of doubt. What if our senses deceived us much more frequently than we think? And, what if, instead of referring

⁸⁶ Assuming the speaker sees furniture in a museum for the first time.

⁸⁷ One may disagree, depending on the characterization one gives to the concept of skepticism. In my view, the presence of positive evidence against the existence of meaning-I gives a *warranted belief* against it and, thus, it is not properly called a skeptical epistemological process. Conversely, the lack of evidence against the existence of objects outside our mind makes this skeptical stance properly defined as such.

⁸⁸ For instance, see Searle (2015). But Wittgenstein himself can be thought to agree with this position, as in his (1969).

to a chair by the word “Chair” for the rest of your life, your use of “Chair” directs to a table from 2030 onwards? Interestingly, if the similarity really held all the way, it seems to me that a meaning indeterminist would be forced to support skepticism about the (correct perception of the) outside world as well: after all, if two claims share the same argumentative defense, and the argument is considered sound in one of them, then the second one also needs to be supported. Fortunately, I find the two positions to be very far apart if looked at more closely, due to one important difference: the sheer *possibility* of having a hallucination should not give us a right to claim that all perceptions are unreliable, because we have no good reason to doubt each of them. Thus, it should be considered rational to trust our perceptions in normal conditions, as long as no evidence is brought forth in support of their unreliability.

Now, I suspect Warren has something similar in mind, in the context of the Kripkensteinian discussion, when he says: “When S has an M-general⁸⁹ disposition to φ in C, given an arbitrary situation w in C, it is rational to be nearly certain that S φ s in w ”⁹⁰. Warren’s argument is very close to the naïve realist’s I just discussed about the external world: as long as no proper reason is put forth to doubt that we are meaning what we (take to) obviously mean by our word –such as a chair when we say “chair”–, then we are rational to claim that we mean *Chair* by “chair”. In other terms, it is simply unreasonable to imagine that our rule should be different than the obvious intuition we have about it tells us⁹¹.

The fundamental problem in the comparison with the unreliability problem about perception is that we know for a fact that perceptions and the outside world exist, and there are no good reasons to think the opposite, thus it is reasonable to regard perceptions as trustworthy. Importantly, as far as I can tell, no good argument has been brought forth to establish that either the perceptual act (as a fact in the world⁹²), the outside world, or the combination of the two, are incoherent in any way.

In this context, on the other hand, the entire endeavor begins from KW’s challenge that meanings *do not* exist (the negative metaphysical thesis), and good reasons to doubt their existence are put forth (such as the compatibility of any word with infinite rules). Now,

⁸⁹ An “M-general” disposition is such that, in the *overwhelming majority* of cases (thus, not necessarily always), it accords with the rule.

⁹⁰ Warren (2020), p. 271.

⁹¹ This sort of idea is a relative to the Simplicity Argument and to Reference Magnetism: they all seem to make indirect use of a plausible anti-skeptical thought process, which doesn’t really work in the context of KW’s challenge, because, if we want to call it a skeptical challenge, it is of an entirely different epistemological nature than the “classic” ones (as explained in Section 1).

⁹² Naturally, many arguments have been brought forth to establish that the concept of perception *is* incoherent, as understood in the “sense data” doctrine. It is rather its existence as a mental fact among other natural facts that dissolves such entanglements.

with a solid metaphysical ground, the epistemic doubt is entirely warranted and, thus, not skeptical in the traditional sense of the term at least.

Moreover, the dispositionalists are supposed to ground meanings in actions: either they accept *all* actions, or their acceptance of *some* actions and not others seems completely arbitrary. Let us apply this observation to Warren's argument. Say I used the utterance "chair" 100 times in my life. 98 of those times, I referred to chairs, while I called "chair" a table twice, maybe due to its unusual shape that confused me. Warren would say that the meaning-S I am following here is *Chair*, for the good reason that the deviant cases constitute only 2% of my usage. But say that, out of those 100 times, I referred to 98 *red* chairs and 2 *red* tables, and those 100 pieces of furniture are all the red objects I ever referred to. By his reasoning, it would be more rational to expect, given these pieces of information, that what I mean by "chair" is *Red object*.

The basic problem Warren can't seem to defend himself against is thus the "Privilege argument"⁹³. In his argument, he accepts the existence of alternative dispositions: there are instances where the speaker uses a word incorrectly. The problem is that he goes on to assume that the dispositions we *expect* to be correct are the correct ones, though the entire point of dispositionalism is to find a way to *justify* the correctness of some dispositions and not others: otherwise, choosing a set of dispositions among the infinitely many different ones is perfectly arbitrary. Now, admittedly, Warren predicts this objection: "Paul Boghossian has offered detailed arguments purporting to establish that there is no way to finitely and non-question beggingly single out a class of situations in which all errors are impossible"⁹⁴. Just after stating this, though, he goes on to argue that "to solve the problem of error, all that really matters is that our dispositions clearly distinguish the 'correct' answer from 'incorrect' answers". Yet, this is precisely what *cannot* be done by a dispositional approach, even in probabilistic terms, as I have already argued. Since "any course of action can be made to accord with a rule", there is in principle no way to establish, without begging the question, which rule our dispositions are following and, importantly, *which dispositions we should count in or rule out*.

In conclusion, Warren's account has more solidity than many other attempts to support dispositionalism, and I don't believe that we could find a proper reply to it within WRPL. But that is precisely because Kripke didn't interpret the algorithmic approach as a *dispositional* approach. If framed in that way, the algorithmic answer falls prey to some of the usual counterarguments, though not all of them. By themselves, to be sure, *Error* and *Finitude* seem to lose some of their strength, due to Warren's resort to the notions of Normalcy and Stability. That said, the algorithmic approach falls prey to the other two fundamental difficulties any dispositional account struggles with: it gives no answer to the

⁹³ As Guardo (forth.) claims, this is the main "Achille's heel" of Warren's theory.

⁹⁴ Warren (2020), p. 271.

hypothesis that, in general, the act of counting marbles is not the product of chance or of a different rule (*Normativity*) and it gives no justification to accept this procedure as the procedure for addition rather than other procedures the counter may be disposed to follow (*Privilege*). Finally, if we are to take the algorithm not as a practice, but as a rule to be followed, then we return to the objection that was fatal to IDA: it is question-begging to explain what the meaning of addition is, by assuming a rule for addition. Hence, the algorithmic response actually survives KW's attacks if interpreted in dispositional terms but doesn't survive all of his attacks to dispositionalism itself.

With these objections in mind, let us now look at Båve's (2019). The title itself of his work, "Semantic Dispositionalism *without Exceptions*"⁹⁵ gives the feeling of a fundamental misunderstanding. As expected, in fact, over the course of his work, Båve looks for plausible ways to prevent the "disparate exceptions" to dispositions that normally work just fine as an analysis for meaning. His worry goes towards the fact that we can't *always* account for our dispositions as adherent to meaning-S. His arguments are plausible within the context of meaning dispositionalism, that is, *assuming* that dispositionalism is the right answer to KW and that, in normal circumstances, we correctly adhere to rules. But we have seen that even if, theoretically, it was possible for our dispositions to always accord with the rule prescribed by meaning-S, dispositions would not capture the *nature* of meanings. So, this approach may be attractive to theorists who are already convinced by dispositionalism, but it really has little appeal for anyone else.

To sum up the most recent contributions to the dispositional cause, the common denominator to support dispositionalism in the first place seems to be a probabilistic perspective: Warren finds it plausible to salvage dispositions despite peripheral exceptions to the rule, while Båve grants the plausibility of dispositionalism as an analysis of meaning and works to prevent such exceptions from being there. Neither author hits the mark, though, because, unlike skeptical scenarios that ask us to doubt the existence or reliability of something that *obviously* exists⁹⁶, KW's challenge gives us solid reasons to think that meaning-I and meaning-S may very well not exist. If the dispositional answer still hopes to find success, then, it needs to frame our linguistic behavior in a way such that it can make sense of the normative nature of meaning, without falling for the -indeed very powerful- *Privilege* objection.

Naturalized dispositions

Hopefully, the discussion above has shown, at least to some extent, that the dispositional account does not seem to be able to save meaning-determinism. Still, I believe this position to capture an important aspect concerning communication. Dispositions are successfully used in the philosophical literature to analyze a different intentional act apart from

⁹⁵ My emphasis.

⁹⁶ Or that we have no good reason to doubt.

meaning-I: desire. I will assume, for the sake of this discussion, that the “most widely-held theory” of desires is the correct one, that is, that “having a desire is a matter of having dispositions to act”⁹⁷. My suggestion to naturalize what dispositionalists analyze as meaning-I is to find the dispositional aspect of communication in the -properly dispositional- desire to produce a behavior in the interlocutor. Let us see a first, basic example:

At the greengrocer’s, I want an apple. This desire can be analyzed in dispositional terms: ‘If at the greengrocer’s + other contextual conditions (having money, being hungry, following a healthy diet...), then: acquire apple’. The greengrocer’s desire to sell me an apple can be translated in the same way. Now, communication makes the coordination of our desires possible, or, at the very least, it is helpful to that end. What happens when I say “Hello, may I have an apple?” is that I make a specific set of sounds that my interlocutor was socially trained to recognize as the expression of a desire. Since their desire is to accommodate mine in this case, they will reply “Of course”, for the same reasons, until both desires (my having an apple; their having sold an apple) are satisfied.

This is a classic example of how communication without meanings can work out⁹⁸. I want to make it clear that, in my view, this account holds with full generality across all kinds of speech acts, even those where singling out a desire to make others behave in some way seems difficult.

First of all, notice that our desires about other people’s behaviors can be “disinterested”, that is, without any aim at our own advantage. We may want to make others laugh, cry, leave us alone, find us friendly, find us menacing, etc. None of these reactions usually lead to a material or even psychological *gain* on our part, unlike the greengrocer’s example: they still are accommodated in my theory, because they are still communicative acts that make people change their behavior through the use of words or other symbols.

Here is another difficulty: sometimes it would seem that all we want to do with our words is to hand over knowledge, such as when we give street directions, or when we write down some notes on a notebook for future reference. This is an interesting objection, because there doesn’t seem to be any change in behavior involved⁹⁹. Nevertheless, I disagree.

While we certainly want to transfer knowledge in such cases, there is always an underlying desire for action on the speaker’s (writer’s, signaler’s...) part, whereas it is not

⁹⁷ Schroeder (2015). Naturally, there are dissenting voices. I can’t address them in full here, but, in a nutshell, I think the other prominent accounts of desire tend to make sense of specific instances of desire, or of what desires ought to be, not of what all desires *actually* have in common.

⁹⁸ An equivalent, the “slab workers” case, can be found as early as Wittgenstein (1953), which was then reutilized more to the point by Guardo (2019).

⁹⁹ Thanks to Professor Andrea Guardo for pointing this out in private conversation.

always the case that communication is intended as a transfer of knowledge. Let us see some further examples, to illustrate the point.

CASE 1:

Paul asks me for directions to the nearest drug store, and I reply: "It's right around the corner!". His speech act has an obvious aim, similar to my asking for an apple at the greengrocer's: to cause me to act in such a way that I give him the information he wants. But *my* speech act is also the expression of a desire, even though it may be more "hidden": I may *want* Paul to act according to his desires, that is, to get to the nearest drug store¹⁰⁰, or maybe I just want him to find me friendly, or to not get angry at me. This point is very clear if we imagine a different state of affairs where I *lie* to him: if I deliberately give him a wrong information, that is because I want him to get lost, or to leave me alone. Either way, it is clear that in this exchange we are using our words to produce actions in the interlocutor, which *in this case* are triggered by knowledge of the world, or lack thereof. Such knowledge is granted by our common psychological attitude towards words and their referents, as we will see in Section 4.

CASE 2:

During a conference, I write down notes on my pad, in order to re-read them later on and have access to the knowledge they convey. What may seem odd about this example is the fact that only one speaker is involved: if a single speaker desires his own behavior to be in such and such way, he doesn't seem to need language in order for that to happen. My take on this is that our *current* selves want our *future* selves to be able to act in a specific way (or set of ways), which is, in this case too, made possible by the acquirement of knowledge. It is not unusual -at least for me- to think whether my notes will be discernible for my future self after months, or even years: that leads me to write them down more carefully and clearly, because the information I will have in the future may be different and, thus, make it harder for me to understand *myself*.

After these two examples about knowledge acquisition, it is important to notice that there are many ways to acquire knowledge of facts in the world. A very common one is through direct *experience*: I can come to know that the drug store is behind the corner by walking around the corner, for instance. Here we have no desires in place: there is no "ought" in the process of learning here, whereas an information given by another speaker can be understood as: "You ought to walk round the corner if you want to reach the nearest drug store".

CASE 3:

¹⁰⁰ More generically, we could say that I want Paul to have a specific *range* of options, that stem from his knowledge of facts in the world. Thus, I may not phenomenologically *want* him to reach the store, but I certainly want him to have the chance to get there, since I give him the information.

At dinner, I talk to my family about my day. They listen to me and, in turn, talk about their day as well. No behavior seems to have been elicited by these words, just a disinterested exchange of information. Another important aspect about this example is that, unlike the previous examples, the exchange of information is usually perfectly gratuitous: crucially, *the content is not the point of the communication*. Intuitively, then, one might say that there is no desire for any elicited behavior here.

To the contrary, I don't believe this last intuition should follow from the context, specifically because it is a mistake to interpret the notion of "behavior" in such a narrow fashion here. A behavior, in the way I frame it in this discussion, is any response whatsoever given by the self or by another party to the communication, be it empirically manifest or "hidden" to the view in our inner lives. With this specification, it should be clear that the family members always have a "behavioral goal", oblivious as they may be of it: they want to receive opinions on what to do with the co-worker they can't stand; they want to boast about their successes and complain about their failures; they want to bond with each other or start an argument. Not only that, each individual in the communicative act can be seen as continuously reinstating their social role within the community they are communicating with¹⁰¹: so, we may say that the communicative act elicits a number of behaviors both in others and in oneself. Each speaker does not need to consciously set out to cause these microscopic elicitation projects: the vast majority of them is entirely subconscious.

In the example, all the communicative acts of each family member correspond to expected *correct*¹⁰² behaviors on the other family members' part: they are expected to lash out at the co-worker, to commend an accomplishment, to laugh together. Precisely because the content doesn't take center stage here, this is a great example to see that their words are used in a way that cares very little for a precise meaning attribution: as long as the right "mood" is set among the speakers, the actual semantic contribution matters very little. At the same time, as we said, the practice of communication is essential for the *self* to settle itself into a social role: the father will communicate (and behave) in a certain manner, maybe trying to assert his authority, maybe trying to look more loving; the son will do the same, depending on his own relational goals; and so on. Of course, this is not to say, to the other extreme, that our words don't matter at all in these kinds of contexts, but there are two points to highlight here:

1. That the words one uses (or doesn't use) *are* a component of one's behavior, just like a certain attitude or a certain gesture;

¹⁰¹ See Watzlawick (1967) on this point.

¹⁰² "Correct" here means, at the very least, that it is commended by the other speakers, similarly to KW's solution suggests. We will see that this usually happens in communication, but it is neither necessary or sufficient for an expression to be correctly used; nevertheless, for now the idea will do, since we are looking at what *the other speakers deem* correct.

2. That sometimes -especially when, like in *CASE 3*, the passing of information is not particularly relevant- the elicited behavior will be elicited *whatever we say*, as long as the content is not too far from the interlocutor's expectations.

It is important to linger for a moment on the point about the speakers' level of awareness concerning their dispositions: only a very limited portion of our speech acts can be said to aim at an "interest" in a strong sense of the term, one that, that is, the speaker is conscious about and actively craves. Examples of that kind are the "Slab" case and the greengrocer's case I introduced earlier, where the speaker's utterance is directly associated to an object of desire (the slab; the apple), and where the other's speaker's behavior (passing the slab; selling the apple) is only a means to that end. Usually, though, we use words and other signals because they influence the interlocutor's (and our own) attitudes, but not because *we know* they do. This, as I will explain when introducing Brian Skyrms's work, is something we have in common with many, if not all, other animals: we behave in ways that make others inclined to act *accordingly* (which is not to say, "in the same way" or "as others would", *pace* Kripkenstein's solution). It would be an unwarranted jump to claim that we need to be self-conscious about it: actually, in the overwhelming majority of cases, we don't know or even think about what effects our words have on others¹⁰³. Essentially, we can say that all we need to support this idea is that it has a biological, and not an epistemological, foundation on the speaker's side.

To sum up, whatever the circumstances, any communicative actor has a (conscious or subconscious) *drive to make something happen*: this is how I believe we should frame the dispositional aspect of communication in psychological terms.

With these specifications, I hope I have cleared up at least some of the reasonable doubts that the reader might have had about the account I am proposing. My central point here is that there is nothing *special in nature* about linguistic communication: words are just the most *sophisticated* device to produce behaviors in others¹⁰⁴, among many: a wink, a shout, a slap, a gunshot, a unique gait, all work in the same way and, unsurprisingly, are usually collected under the label of "nonverbal communication". What these acts may be understood as lacking is, tautologically enough, the "verbal" component, and the one non-

¹⁰³ This contradicts Lewis (1969), an author we will discuss later on, as he deems the awareness of the speaker's and the interlocutor's intentions to be essential for successful communication. I take it that this is simply empirically false, but I will come back to the point in Section 4.

¹⁰⁴ It could also be noticed that, sometimes, we consciously use words to produce actions in *our present selves*: an example could be to whisper: "Come on, you can do it!" before walking on stage in front of an audience, or the habit some people develop of talking to themselves. I leave the specific reasons for this kind of behavior to psychological studies, but I suspect a game of "make believe" may be at play here, which introduces a fictional interlocutor to help us get through an emotionally stressful or cognitively challenging situation (see for instance Kirkham 2012 for an interesting study on how speaking to oneself consistently improves the success of goal-directed behaviors): this is all the more reason to support the idea that communication has a powerful influence on our psychological lives and, more specifically, on our resulting behavior.

problematic difference between the two forms of communication appears to be the presence or absence of words in the endeavor. Whether that implies that verbal communication brings something *more* to the table on a semantic level than the non-verbal one, remains to be seen and is, in my view, false.

Thus, communicative acts are, at their core, *conveyors* of behavioral desires and, as such, of dispositions: the subtext of what I tell the storekeeper is: “Given the situation, I am inclined to have an apple”, which in turn implies “I desire for you to sell me an apple”. We have seen a few other examples, and the reasons to think that even when desires are subconscious and/or not immediately directed to an object of our craving, successful communication always leads towards a more favorable interpersonal context from the utterer’s perspective¹⁰⁵.

Importantly, there is nothing we can fail to comply to, here: we are not following a rule when we desire something, because desires *really are* attitudes towards something, whether it exists or not and whether it is right or wrong to. It is possible to desire a six-legged dog, even though six-legged dogs don’t exist; it is possible to desire to kill an innocent person, even though it is immoral. Desires are what they are, it is not a part of their nature to comply to anything outside of themselves when they arise¹⁰⁶: that is why dispositions work so well to define them, and why they don’t work to define meaning-I, which is, instead, exactly defined as an act of compliance to a rule. Notice that I am *not* claiming that communication is *just* a conveyor of desires: as the examples above have shown, it is also, for instance, a conveyor of information, or a pastime. What I am claiming is that the *dispositional* aspect of communication, which dispositionalists wrongly interpret as semantic, is to be found in, and explained in terms of, that component of our linguistic practices.

I think there is good reason to assume, then, that in analyzing communication we should focus our attention on its nature as an act among other acts. As such, desires are an essential aspect of its realization, whereas it is far from clear that meaning-I should have such importance. To make my claim clearer: I am saying that meaning-dispositions are nothing over and above desire-dispositions, or, if you will, that the first don’t exist and the second do; thus, desire-dispositions are helpful in explaining how communication works, whereas meaning-I can be left behind. Thus, when dispositionalists concern themselves with meaning dispositionalism, they really ought to refer to desire-dispositions concerning our communicative acts.

¹⁰⁵ This is not to say that the context *is* objectively favorable for the utterer, but that the utterer’s point of view is that their context will be more favorable. The objective aspect of the interaction is the effectiveness, or lack thereof, of their communicative act.

¹⁰⁶ Obviously, there is another sense in which desires require a compliance with facts, that is, in their *satisfaction*. But here we are only considering their arousal, as it is the aspect we are assimilating to communicative dispositions.

Another conclusion we can sketch here¹⁰⁷ is that the sounds we utter and the words we write are not “lassos” we use to catch meanings-S, but a natural part of our behavior with a functional role in our psychological lives, just like a shrug, a sad face or a scream are, though at a lower level of sophistication.

I have stated that Båve (2019) and Warren (2020) misunderstand KW’s problem with dispositions. They interpret KW’s challenge as posing a problem to the dispositional account because of “exceptions” to the rule, whereas the real problem is that *no disposition* whatsoever works as a proper analysis of meaning-I. As we have seen with the Privilege Argument, in fact, the supposed exceptions, mistakes and errors may very well be part of the rule we are actually following.

That said, I do think that this confusion tells us something important, though. When we communicate, our dispositions (desires) are usually satisfied: *exceptionally*, this doesn’t happen. It doesn’t happen when we can’t coordinate because the interlocutor can’t or won’t behave in the way we expect our words to make them behave. Say, for instance, that you are practicing for a mathematics test with a friend. If your friend asks you to calculate the sum of two numbers too big to even write them down over the course of a lifetime¹⁰⁸, you *do not desire to give the arithmetical answer*. You may very well desire to leave the company of your friend, or to laugh the question off. This is what KW, in the context of the refutation of dispositionalism as an appropriate answer to his challenge, called *Finitude*; translated in terms of desires, it turns into something very intuitive: we don’t always want to accord with the behavior our interlocutor expects from us¹⁰⁹.

Some other times, you would, in principle, want to accord with mathematical addition, but your cognitive abilities prevent that from happening: you make a mistake, thus keeping your disposition from being satisfied. This is the psychological equivalent of *Error*, but instead of being a reason to reject dispositionalism, as the “logical” version does, here the proneness to make mistakes is accounted for. Notice that we are now legitimized in reasoning this way, because we are not trying to connect to the entire function *Addition*, but instead we want to accord with a fraction of it and we may or may not be able to. This is to say that dispositions *can* account for the meaning of a word in a very loose sense, quite different from the one meaning determinists support: it is explicated in terms of the “direction” we want our collective attention to go towards in the world (on this topic, see the notion of triangulation in Section 4).

Now that we have turned dispositionalism into a theory about our *desires* in communication and not about our meanings-I, the “exceptions” Båve talks about make

¹⁰⁷ But I will focus on it in Section 4.

¹⁰⁸ I am again referencing the famous example in WRPL, pp. 26-27.

¹⁰⁹ The “extreme” setting of this example may make this harder to grasp than it really is. *Finitude* works in everyday life contexts as well: say you are not disposed to answer to “2+2” because you are tired after a long day at work, or because you didn’t hear the question properly.

perfect sense: only sometimes, and definitely not the majority of times, our words do not get the result we would like them to have on others. Similarly, when Warren argues that we have reason to be fairly certain that someone means something because, even if there are mistakes, they are rare and far apart, he really says that we can reasonably expect a kind of behavior from someone when they *usually* accord with it: if I want Hannah to focus her attention on a pencil, I will say “Pencil!” and, due to our common nature and nurture, she will *most likely* and *in most occasions* focus on the pencil.

I hope I have provided a sufficiently plausible psychological narrative for “simple” dispositionalism in naturalized terms. An interesting question would now be whether there is place for a psychological understanding of the notion of *idealized* dispositions as well. This would require a way to psychologize the notion of idealization, which is not necessarily a simple task. That said, I think there may be a solution to this further riddle, and I will try to present it in terms of “capability dispositions”.

Capability dispositions

Call “capability disposition” a desire-disposition about the speaker’s ability to do something¹¹⁰. By “ability” I don’t simply refer to one’s physical or cognitive skills: as does the dispositionalist, I refer to the totality of conditions that make the intended action possible. Say, for instance, that my friends mock me for being bad at basketball: this may produce in me the desire to be good at basketball and thus the disposition to change my relevant skills (aim, endurance...) for the better. But not only that: I will also desire external conditions that make it possible for me to utilize my skills for the better (good weather, spare time...).

Crucially, this kind of desires are partially responsible for the more or less frequent satisfaction of my “first order” desires: if I want *to be able to* score a basket, it is practically obvious that I *want* to score a basket and it will usually be more likely that I *will* score a basket (because I may be induced to train myself to that end, for instance). Thus, we can say that the desire to have an ability is itself one of the conditions for the ability to be there, though neither a necessary nor a sufficient one.

Now, in the context of communication, idealized dispositions are capability dispositions about my making the interlocutor behave in some specific ways: I want to be able to communicate in a way such that it produces the desired behavioral reaction, which basically implies that I would want to communicate in said way, were the conditions

¹¹⁰ One may be tempted to call this a meta-disposition: a disposition (desire) about a disposition (ability). I find this terminology problematic, because I don’t think abilities can be analyzed with the same kind of dispositions as desires, or at the very least their relation is quite complex. One observation we may want to stress is that one of the conditions for a desire-disposition to be actualized is that the subject must have the ability to make it happen and that one of the conditions for an ability-disposition to happen is that the subject must have a desire (in a loose sense) to make it happen. I am not going to investigate this relationship further here, and it doesn’t seem particularly important for my point.

favorable. Again, my desire for the ability to be there is a valuable help, but we should not consider it as necessary in view of the communicative result: speakers normally communicate without desiring to be able to do it; they just have the first-order desire-disposition to make others behave in some way.

Say you want to make your friends happy by giving an answer to their overlong mathematical question: even though you probably don't really want to answer, maybe because the addenda are too long, or because you have a headache, or because you fear you are not smart enough, you may *want to be able to do it* (you may desire to have the patience, the cognitive stamina, the time to do it). This is what an idealized disposition, in psychological terms, comes down to: "If I had the time/if I was smart enough/if I didn't have a blindfold on/etc., I *would want* to say *x* because of the behavioral reaction it would imply. In reality I don't want to do that, because I don't have the time/I am not smart enough/I am physically impeded..." and thus I cannot make the behavioral reaction happen. I would like to stress how the *ideal* nature of idealized dispositions is conveyed here. The reader will notice that there are two sides to it: one is functional, the other strictly psychological. The functional aspect is fact-based: for the event to happen, this list of facts need to occur, that is, this is the ideal condition for the further fact to occur; the psychological aspect is related to the speaker's desiring for the fact to happen, that is, the subject's ideal state of affairs is such and such in his mind. Both aspects seem to make sense of two aspects of dispositionalism: respectively, the problem in metasemantics and the one in the philosophy of mind¹¹¹.

Again, it is important to point out why this version of idealized dispositions does not fall prey to the objections we saw earlier. While a classic dispositional theory will argue that the speakers' idealized disposition is to say *x* because it is what they mean, falling for the usual vicious circle, the naturalized dispositional theory argues that the speakers' idealized disposition is to say *x* because it is what will produce the desired response in the interlocutor. Since the ideal outcome lies *outside* the realm of semantics, it is tenable. Moreover, shifting from the realm of logically apt answers to the realm of psychologically apt answers frees us from being bound to *exact* rule-following and, thus, from the strong notions of meaning-I and meaning-S we have been struggling to come to terms with. There may, in fact, be countless different linguistic behaviors that produce a behavior that is deemed as desirable by the speaker; the reference to a precise meaning-S in communication does not allow for that.

Let us now see if capability dispositions help make sense of Båve's and Warren's talk of "exceptions". As we said, both philosophers assume that dispositions work in most normal cases: the exceptions are taken to be deviations from the norm. Warren even goes

¹¹¹ Again, this is not to say that this is a direct reply to those problems. In my view, meaning determinist dispositionalism is misguided for that reason: because its supporters confuse these features of communicative desire-dispositions with features about meaning-I and meaning-S.

on to argue that all it takes to be justified in believing we are following meaning-S *Chair* is that we follow that rule “usually enough”: so, the museum example is a physiological mistake, an “exception that proves the rule”. *Privilege* has blocked the way to success for this theory, because there is no logically stringent way to decide which past behaviors should be taken to be the norm and which aren’t. But this objection has no power in the psychological framework: here we do have a reason to choose some behaviors over others, because we are not strictly following any logical rule, we are *environmentally influenced* to prefer one answer over another, or, as we will call this phenomenon in Section 4, our environment leads us to *salient equilibria*.

Due to all this, I claim that the dispositional answer is not a tenable position in its standard, meaning determinist, arrangement, but also that it makes perfect sense in a psychological perspective and it can be integrated in a wider, meaning indeterminist account as one (relevant) aspect of human communication.

Conclusions

We have reviewed a number of famous and influential dispositionalist accounts spanning from the ‘80s to the current state of the art. None of them have given sufficient arguments to save this theory as a meaning determinist claim, essentially due to the impossibility to overcome either *Naïve Normativity*, *Privilege*, or both. What the future literature on dispositionalism should take as its top priority, if it wants to defend meaning determinism, seems to be two-fold: on the one hand find a way to explain why we mean anything by what we are disposed to say in actual or ideal conditions; on the other, explain why we should be justified to believe we mean one meaning-S rather than any other. This “preference” may be explainable in terms equivalent to Lewis’s (1983) defense of the existence of “natural” universals; but, just like in Lewis’s theory, there doesn’t seem to be a compelling reason to think that what *we find natural* is really what exists in the outside world, *if* we have good reasons to doubt its existence in the first place as a class of objects.

We have also tried to make sense of dispositionalism by changing its purpose: instead of focusing on meaning in communication, dispositions should be an analysis of what we *want* to convey with our words in our interlocutor. This has been made possible by another shift in perspective: instead of looking for a logically faultless rule, we have channeled our communication practices in a psychological environment, giving center stage to behaviors and the reactions they produce.

This, in my view, is enough to answer one difficult question for meaning indeterminism: how do you make sense of the “internal feeling” speakers have when they, supposedly, mean something? Actually, many of the views that take as a necessity to save meaning-I do it because of this: they defend a Moorean-like feeling that we seemingly can’t get rid of in our theories.

Kripke considers this problem as well, but concludes that even if there was such a thing as an “experience of meaning”, there would still be no way to *logically* determine which meaning is the one we are experiencing. Phenomenologically speaking, though, we all *feel* like we never meant *Chair-2030* in our lives and yet we did mean rule *Chair* when presented with one. How do we make sense of this?

I think we have presented a partial¹¹² answer to this question. An aspect of that feeling is a desire to make others behave in a specific way or get in a specific mood. Unlike meaning-I, desires are unproblematically considered intentional acts, because, as we will see better when discussing Davidson’s triangulation, they are directed to something *in the world* and to something that usually exists, whereas meaning-I is supposedly directed to a category of objects that, by their very nature, probably doesn’t exist. Ironically, then, even though dispositionalists are concerned with the metasemantic question, the new reading we gave to their proposal seems to satisfy some of the conditions to answer the question in the philosophy of mind.

Where does meaning-I fall in this picture, then? As an anti-dispositionalist about meaning and a meaning-indeterminist, I say *nowhere*. There is no need for a mental state that adds nothing to the descriptive picture of how communication works, and that has turned out to be – thanks to KW’s contribution – extremely problematic on its own terms. Unless either it is shown that the picture lacks something essential, or that meaning-I can be made sense of in an easy way, removing it from our ontology seems a more than welcome philosophical move. Primitivists try to argue for the first of these two claims, and give the easiest way possible to reintroduce meaning-I from the back door. Let us see if they succeed.

¹¹² It is reasonable to expect that something more must be included in the phenomenology of “meaning something by one’s words”. For instance, it seems inevitable to add a sensation of “grasping the rule” one feels he’s referring to, as Kripke (1982) would say. What really matters for our discussion is that, whatever the correct description of this internal feeling is, “we experience language artifacts and gestures to have meaning-properties *even though they don’t*” (Azzouni 2013, pp. 350-351, my emphasis).

3: Primitivism

After addressing the most popular reply to KW, dispositionalism, I stated that we can do without meaning-I better than we can do with it and that, because of this, meaning indeterminism gains credibility as a good candidate to be the best answer to KW's question. This is not a popular opinion, though. If I managed to show why the dispositionalist project seems doomed to fail, then we have dramatically thinned out the meaning determinist ranks, but that is not to say by any means that there are no other interesting options to consider. There are two main alternatives to dispositionalism and meaning indeterminism in the literature that are worth considering: one, primitivism, will be discussed in this section. The other, reference magnetism, will not be addressed in this essay, although I am confident further work can uncover its significance in the naturalistic account I am proposing.

One prominent reason why some non-dispositionalist philosophers prefer to choose another meaning determinist thesis instead of meaning indeterminism is that KW seems to find himself in a tough spot still: in particular, it is still not clear how to make sense of the distinction between *correct* and *incorrect* applications of a word, if there is no proper reference to a general rule such as meaning-S. Granted, an important part of KW's own solution is concerned with deflating the concept of correctness conditions in favor of assertability conditions, but this doesn't really satisfy the intuitive need to find a way to determine whether *anyone* is right or wrong in using a certain word in a certain way. Correctness conditions are, by definition, able to determine whether an assertion is true or false; assertability conditions merely determine whether, in a specific context, speakers are disposed to use a certain assertion. But, the meaning determinist will say, why should we care about the speaker's proneness to use an assertion instead of looking for the grounds for its *correct* use? We will see later that I don't find KW's solution valid either in terms of correctness or on his own. I claim that is a fundamental reason to support meaning determinism and, in my view, the best one. In other words, meaning determinism would be the best theory of meaning we have if the only meaning indeterminist account available was the one found in WRPL. In fact, it seems to me that if KW is happy with doing without the correctness/incorrectness distinction, he is forced to relinquish communication as we know it, and to frame it as a completely, or at least heavily, "blind" practice without the possibility of proper "mistakes" or of the conveyance of exact concepts from one speaker to the other.

As KW is cornered in this tight spot, the Primitivist answer gains some of the credibility that it may lack at first glance. As we have sketched out in the first Section, at its core there is a very simple, yet bewildering, idea: meaning-I must exist, KW shows that we have no good way to explain it, thus we must take it as primitive, that is, as a fundamental brick upon which to build everything else, in metaphysical terms, or upon which to build one's

set of beliefs, in epistemological terms. In the eyes of the primitivist, KW's statement is not just, as a dispositionalist may say, false, but it is fundamentally self-undermining, because the sentence "There can be no fact as to what I mean by any word, at any time" must already have a meaning-I connection to even make sense or be understandable by anyone.

This specification may seem redundant: after all, Kripke has called KW's claim a paradox since WRPL itself. There seems to be a fundamental difference here, though. Kripke calls KW's suggestion a "paradox" because it puts two very common intuitions against each other: this creates a conundrum difficult to untangle, since, at first glance, we don't seem to be able to reject either of them. The Primitivists, on the other hand, see this as an untenable position, because they assume meaning-I to be simply ineliminable *a priori*: it is not a matter of competing intuitions, but rather a matter of the precedence we need to accord to meaning-I, which KW doesn't accord when he supposes we may do without it.

It is in fact a capital question about this "fundamentality" of meaning whether we are talking about an epistemological or a metaphysical kind of fundamentality. I understand much of the literature on primitivism to start from an epistemological perspective and then conclude that the primitivity of meaning is metaphysically inevitable: we cannot *know* or *understand* meaning-I in terms other than itself because in order to know or understand something we already need to mean something by the words that make up the sentence concerning meaning-I that we claim to know or understand. On the side, this may have consequences in the philosophy of mind as well: thus, we may claim that to be in the intentional state of knowing or believing that *p*, we need to be in the intentional state of *meaning p* first and foremost. This version of primitivism seems to me *prima facie* intuitively plausible.

If meaning-I is to be taken as also metaphysically fundamental, then that means that meanings exist in the world without any other objects as their grounds. If the reader is familiar with David Chalmers's work, it will be clear that this is the equivalent of his fundamental phenomenal facts, but in the realm of semantics. To my knowledge, this position has been defended only very recently for the first time, as it doesn't survive the "intuitiveness check" quite as well. That said, the arguments in its defense are at the very least interesting, so we will look into them.

Meaning determinist primitivism and meaning indeterminist primitivism

Another important point is that the Primitivist Answer I will be challenging in this section is of the meaning determinist type: the claim that meaning exists mind-independently and that it cannot be reduced to, or explained in terms of, anything else. It is very important to understand this point, because we will see, in the final part of the essay, that the evolutionary meaning-indeterminist answer can be alternatively understood as a

primitivist answer in psychological terms¹¹³: in this case, the claim is that, *for reasons that are not logically justified*, but rather dependent on the environment of the parties to the communication practices, language is the way it is and not any other. One crucial difference between the two accounts is that meaning deterministic primitivism takes meanings as determined *a priori* and as defining the correctness of use of a word by a logical rule, a function that assigns, for every context, the proper response; the account I will defend is, instead, a *naturalistic* explanation for our use of a certain word in a certain context: it is both empirically grounded (having an *a posteriori* reason) and entirely contingent on the actual state of affairs.

The two views also diverge under the “primitivist” aspect. Meaning-determinist primitivism¹¹⁴ is the claim that meaning-I must be taken for granted without further analysis, usually justifying this claim by pointing out that meaning-I is an *a priori* condition for any analysis or argument. This is to say that we literally ought not to think about meaning-I’s role in our linguistic endeavors because it is a *precondition* of our even having any content to think about. The sense in which a meaning indeterminist usually speaks of “primitivity” is not of this kind; it is rather an attempt to convey the idea that communicative practices stem from something entirely *alien* to them, not from another semantic fact, such as the pre-existence of meaning. Because of that, I will tend to call the latter “naturalistic”, “evolutionary” or “psychological” view, to avoid a further terminological confusion.

Supervenience and the Soames fallacy

Over the years, a number of authors have claimed to defend a primitivist account of meaning. Many of them, though, can be more plausibly seen as holding other views, which, ironically, tend to be of the meaning indeterminist kind¹¹⁵. The arguments I will focus on, then, will be concerned with a newer approach, mainly due to Kearns & Magidor (2012) and further explored by Anandi Hattiangadi, which largely employs the notion of supervenience. Because of that, I will briefly introduce the concept and show how useful it is to clarify some aspects in the debate about KW.

A supervenience claim, as concisely and elegantly stated by David Lewis (1983), is “a denial of independent variation”. A less elegant but more exhaustive definition would be that x supervenes on y when, for every possible state of affairs, x cannot change without y

¹¹³ What Kusch (2006) calls “meaning-*skeptical* primitivism”, as opposed to “meaning determinist primitivism”.

¹¹⁴ Henceforth, “primitivism” and its supporters “primitivists”.

¹¹⁵ For a powerful defense of this statement, see Kusch (2006), Chapter 7. Kusch shows how some of the most prominent self-declared “primitivists” about meaning-I are really defending something very close to what KW defends. That, Kusch argues, is because meaning indeterminism is also a form of primitivism about meaning, but in a different way. That way, I claim, is psychological, or natural, rather than logical (see the “Conclusions” from Section 1 to be reminded of what the difference consists in).

also changing. For instance, a physicalist may claim that there can never be a change in mental facts (mind) without a change in physical facts (brain). This means that the mental supervenes on the physical: there is no possible state of affairs where the mental changes and the physical remains the same. Be reminded that, by itself, supervenience says nothing of the deeper metaphysical fundamentality of one of the elements in the relation: all it says is that there is an indissoluble *connection* between the two, so much so that one of the two elements may be inessential to our ontology¹¹⁶. In the mental-physical example, for instance, it is common to argue that, if the mental supervenes on the physical, then the physical must be the more fundamental brick of reality, since many physical objects don't seem to have any mental properties, while all mental properties seem explainable in physical terms. Conversely, say a computer scientist builds two programs that work in perfect synchronization: any input given to one of them will also be instantly received by the other. Here we still have a case of supervenience, but neither program is more fundamental than the other: they both, symmetrically, depend on the hardware they are installed on and on the computer scientist that created them. This is all to say that accepting a supervenience thesis does not automatically amount to accepting that one of the *relata* is more or less fundamental than the other: as we said in the introduction, it is a relational claim, not a hierarchical one.

Another important distinction we ought to consider is between a *global* and a *local* supervenience thesis. Local supervenience has objects as its *relata*; global supervenience has *worlds* as its *relata*. Thus, the computer programs example is a case of local supervenience, since it relates two individual objects; on the contrary, as Chalmers (p.34) observes, biological fitness is a good example of global supervenience that is not also local: "Any world physically identical to ours would also be biologically identical [but] two physically identical organisms can arguably differ in certain biological characteristics [...], one might be *fitter* than the other, due to differences in their environmental context". This seems important for our purposes when considering the two aspects of meaning we have been distinguishing: meaning-I is supposed to be an individual intentional act in the mind, thus it seems to be analyzable in terms of local supervenience; meaning-S, on the contrary, is supposed to be an object every meaning-I in a world refers to, so it should be best analyzed in terms of global supervenience: for instance, a dispositionalist may argue that, given two worlds, if the idealized dispositions of the speakers are the same, then the meaning-S is the same.

Let us now see how supervenience can be a vital tool to help us understand KW's problem better, by briefly looking at yet another defense of dispositionalism due to Soames (1998). Soames claims that there is a fallacy in KW's line of reasoning because he doesn't distinguish between "*a priori* determination" and "metaphysical determination". In his

¹¹⁶ In this sense, the physicalist may want to say that, since the mental can only change if there is a physical change, then mental facts can be "explained away" by physical facts.

view, KW is right when it comes to dispositions not being able to determine meaning¹¹⁷ *a priori*, but the same cannot be said for metaphysical determination, and he argues in this direction through a supervenience claim: it is not at all obvious that, “once we cast our net wide enough among the non-intentional facts [...] metaphysical determination or necessary consequence really fail”. That is: whatever “the smallest set of non-intentional facts about me in the actual world on which meaning facts supervene, I am confident that they do supervene” (Soames 1998, p. 229). What Soames is stating is, in a nutshell, that we may not be able to give an exact account of what physical and mental properties ground¹¹⁸ meaning, but also that we can be fairly certain that they do. In a sense, this proposal is very interesting: it touches on the intuitive idea that semantic facts are very limited compared to physical and mental ones, and on the plausible fact that they are not, *pace* primitivists, fundamental bricks of reality.

Here comes a big problem, though: the meaning indeterminist is claiming that we can easily do without the entire notion of meaning, at least as traditionally understood. Thus, of course we can agree that meaning-S can have no change without changes in further facts in the world, *if meaning-S does not exist at all*. The point here is that supervenience doesn't seem to be a good analysis of something when that something is argued to be non-existent by the opposing party: it is important to always keep in mind that the discussion here is whether or not we need meaning-I and meaning-S in our ontology as traditionally understood, not whether or not they supervene on anything else, because that already implies their existence.

Touching on Soames's theory seems to have given reason to think that supervenience is actually *not* a good device to defend meaning determinism, despite what I said earlier. But that is not the case for the primitivists, who turn the argument on its head. Their strategy here consists in saying the opposite of what Soames tried to argue: that meaning does *not* supervene on anything else because it is fundamental, just like fundamental particles for the physical world. As I sketched in Section 1, then, the primitivist challenge -at least the one I will tackle and that found space in the contemporary debate- to meaning indeterminism is very different from the dispositionalist one: here it is granted that there is no way to analyze meanings, but also that meanings are essential to communication and, thus, they must exist fundamentally, that is, they must be primitive.

¹¹⁷ Soames implicitly distinguished between meaning-I and meaning-S when he makes the distinction between determinations, but the two senses get frequently mixed up. When he talks about “*a priori* determination” meaning-I is concerned; when he talks about “metaphysical determination”, meaning-S is. But then, he talks about “*non-intentional* facts” as opposed to meaning-S, which is non-intentional itself.

¹¹⁸ I talk of grounding here for the same reason why I argued that a physicalist claims that mental facts not only supervene, but are grounded on, physical facts.

Interestingly, primitivists tend to focus their line of defense on meaning-I¹¹⁹, which makes them even more apart from dispositionalists. The reason for this is the supposed *immediate knowledge* of the intentional act of meaning something by my words, whereas, as we saw, there is an inferential gap between meaning-S and the facts that should explain it. This should already make us wary of the primitivist endeavor: there seems to be, in fact, some level of confusion on the relationship between metaphysical and epistemological fundamentality. It strikes me as clear that the fundamentality of meaning-I can only be accepted in epistemological terms, that is, as stating that there can be no epistemic state without a pre-existing meaning attribution to the proposition we believe/trust/know/etc. With a slogan, we might say that we can know nothing, if we don't know what it means. On the contrary, the fundamentality of meaning-S seems to be a statement about metaphysical fundamentality: with a slogan, among the basic ingredients of reality, there are meaning-S facts.

So, the basic claim of fundamentality may be best understood as about meaning-I or about meaning-S depending on the overarching structure of the primitivist's argument: the two prominent examples we will look at are emblematic of these two possible approaches. In fact, although neither is explicit about it, Kearns & Magidor seem more interested in defending the metaphysical fundamentality of meaning-S, from which the possibility of meaning-I derives; Hattiangadi, on the other hand, looks more concerned with the primacy of meaning-I, which she deems necessary to even have an argument on the topic itself: in her view, our priority is to show that meaning-I is primitive, which would *then* allow us to discuss the nature of meaning-S.

The primacy of meaning

It is now time to discuss in depth the two most prominent, fully meaning determinist, defenders in the contemporary literature¹²⁰ that argue in favor of primitivism. The first explicit proponents of the "Semantic Sovereignty" thesis, and of the supervenience approach, have been Kearns & Magidor in their (2012) of the same name. I find the most interesting work in the same primitivist direction, but with crucial and, literally, fundamental differences, to have been developed by Anandi Hattiangadi, especially in her recent (2019) contribution. By tackling these proposals and answering them, I believe I will have given a sufficiently satisfying answer to the primitivist claim and to the problems they see in meaning indeterminism. This is not because there can be nothing else to be said about primitivism, but rather because the cogent reasons for rejecting these proposals touch on the essential nature of the proposals themselves.

¹¹⁹ Though, as usual, the distinction is not always clear and/or even accounted for. Kearns & Magidor, as we will see, are a good example of this.

¹²⁰ Especially in recent years, the literature in defense of semantic primitivism has lost relevance. Still, I believe that one important reason for that is that primitivism has been usually framed in a wrong way. It is reasonable to expect a partial resurgence of primitivism due to this new approach.

With Kearns & Magidor, a clarification is in order from the very start. In fact, as said, the very introduction gives a somewhat confusing picture of what exactly will be the focus of their theory.

The authors assume that intentional facts are semantic facts and that, even if someone may disagree with this identification, their theory still works for both kinds of facts (p. 323). They also produce a number of examples to illustrate the kinds of facts they are interested in defending as fundamental: two of them, representative of all the others, are “‘Water’ means water” and “Jill’s current thought is about coffee”¹²¹. Since, arguably due to length restrictions, they don’t give a detailed argument for their claim in their work, but rather they take it as a plausible assumption, I will take it that it is wrong. As we have seen in Section One, in fact, there are two problems with it: one, that what they call “semantic facts” are, if they exist, very different among themselves (those in the mind, meanings-I, and those in the world, meanings-S); two, that meaning-I, if it exists, is a kind of intentional fact, not the other way around. Of course, since this is my word against theirs, it is not enough to simply deny their statement. Thus, I feel urged to give further reasons, on top of what I already said earlier, not only to argue that the notion is usually used in this way in the literature, but to argue that it should be used in such a way.

As stated, intentional (f)acts¹²² are normally understood in philosophy of mind as “the power of minds and mental states to be about, to represent, or to stand for, things, properties and states of affairs”¹²³. This is a family of facts meaning-I seems to be correctly ascribed to: it is a mental act that is simultaneously directed to an interlocutor and to an object (what we will call later on, “tranguation”); when we say that “Mela” means “Apple” in Italian, we are directing our minds towards two words (“mela”, “apple”), one meaning-S (*Apple*), and the one kind of object it refers to (apples). Conversely, many intentional acts have nothing to do with reference, words, and/or meanings-S: when we fear something, think about something, see or hear something, we are directed *immediately* to the object of intentionality, that is, without any form of symbolic intermediary. Against this, Kearns & Magidor point out that “it is not uncommon to assume that speakers can be in certain intentional mental states such as belief in virtue of accepting certain sentences (either in their spoken language or in the *language of thought*)” (my emphasis). What they seem to imply by this is that, when we have intentional states such as fear of something, or belief concerning something, we already have in mind the

¹²¹ All other examples the two authors give are a re-statement of the first one, that is, other examples of meaning-S facts. It must be said that none of their examples are meaning-I facts, but we can assume they count as semantic facts for them too, since they go as far as to overlap them with intentional ones.

¹²² Intentionality implies an action; thus, all intentional facts are intentional acts.

¹²³ Jacob (2019). Very similar definitions can be found throughout the literature: see, for instance, Crane (2001), Searle (2015).

meaning of the sentence that results from making the fear or belief explicit in propositional terms.

I reply to this by arguing that the fact that we *can* put into words our intentional acts does not make them semantic in nature, nor dependent on semantic acts: this is very similar to arguing that, since we can describe a scene in the countryside, the scene is in some way dependent on the description we write down. We don't *need* to produce a semantic statement such as "I see a dog" every time we see a dog, neither verbally, nor mentally: otherwise, we would need a long time to gaze upon a complex pattern, and animals and newborn children couldn't be said to see things¹²⁴, which is beyond unreasonable. In other words, intentional acts do not imply the need for a "language of thought" or a language *tout court*, at least not if this mysterious concept is seen in strictly meaning determinist terms.

I take it we have compellingly rejected the idea that intentional facts, as they are usually understood in the literature, are semantic in nature. From here on, then, I will discuss the primitivist argument solely for (properly) semantic facts, that is, facts about meaning-I and facts about meaning-S. I am assuming this will not compromise my evaluation of the arguments, as I don't see how it would.

Kearns & Magidor's first argument in defense of metaphysical semantic sovereignty strongly resembles Chalmers's analogous attempt to defend the irreducibility of phenomenal facts to physical facts in his seminal (1996). In *The Conscious Mind*, in fact, Chalmers argues that there are possible states of affairs such that, all physical facts being equal, there are no conscious experiences (the "zombie world", where people exteriorly behave exactly like us and are physically identical to us, but without internal experiences of any kind¹²⁵). If such state of affairs is possible, he concludes, "when God [so to speak] created the [actual] world, after ensuring that the physical facts held, *he had more work to do*" (p. 124). The defenders of semantic sovereignty resort to a similarly fantastical idea: the possible existence of "non-physical ghosts". In this case, the counterfactual creature taken into account is incorporeal, but it can communicate with its fellow spirits due to the existence of semantic facts. The idea is, then, that since it is metaphysically possible for ghosts to exist and for them to communicate in some ghostly language, it is then proved that semantic facts do not supervene on physical ones: after determining the physical facts (and, as we will shortly see, the "spiritual" ones in this world), God had to put in some more work in order to create semantic facts.

This first argument has a couple of important problems, which make it, in my eyes, untenable. Firstly, we need to address the problem concerning the *possibility* of the existence of non-physical ghosts. Could such creatures really exist in some counterfactual

¹²⁴ Plausibly assuming that animals and newborn children cannot produce meaningful sentences.

¹²⁵ In Chalmers's lingo: "There is nothing it is like being a zombie" (1996, p. 95).

context? This closely parallels a popular counterargument to Chalmers's zombie world: who is to say that zombies *can* actually exist? Assuming that they can be precisely conceived of¹²⁶, what if they are merely conceivable, but not possible? I think Chalmers has a good defense, on epistemological grounds, against this accusation. He argues that, unless there is compelling evidence against it, there is no warranted reason to determine that something is conceivable but not metaphysically possible. Moreover, "if some worlds are logically possible, but metaphysically impossible, it seems that we could never know it, [since] the information is not available *a priori*, and *a posteriori* information only tells us about *our* world" (1996, pp. 137-138). I find this suggestion interesting, but one could reply that, if we are not justified in doubting that a conceivable world is also metaphysically possible, then we could also say that we are not justified in thinking that worlds with radically different *physical* laws are metaphysically possible. Fortunately for Chalmers, this is not too big of a problem for his claim: after all, what makes us sure that it will be impossible for an individual physically identical to us, but without consciousness, to be created in a laboratory millennia from now, in the actual world?

That being said, the "Haunted World" proposed by Kearns & Magidor doesn't seem to hold its ground as well. In this case we are not asked to imagine a physical support without phenomenal experiences, but the other way around; the semantic facts are instantiated in something incorporeal, something that lacks what it is reasonable to assume, based on scientific evidence, to be needed for any intentional act to be instantiated: a nervous system. In other words, the strength of Chalmers' argument is that he is *subtracting* phenomenal experiences (something *we know for sure to exist*, at the very least in the first person experience) from the actual world, leaving us with a world of creatures similar to androids. The Haunted World hypothesis, instead, asks us to concede that creatures radically different from anything we have ever come across may exist and, moreover, act semantically. To say it in literary terms, the first theory asks us to imagine a science-fiction scenario that, for bewildering as it may be, doesn't seem to be metaphysically impossible or even impossible in some distant future given the actual laws of physics; the second asks us to imagine a fantasy world, which some may find hard to conceive of, and many may find hard to believe possible.

Another substantial problem for the Haunted World hypothesis is that the "creation of the world" metaphor does not work here: Chalmers is suggesting the possibility of an entirely physical world, physically identical to ours, but without consciousness, his conclusion being that consciousness is something over and above the physical world and thus that the actual world cannot be described in such terms, because in actuality God had to add the phenomenal ingredient to the recipe of the world. Kearns & Magidor, on the other hand, are suggesting the existence of physical *and* "immaterial" objects: but what are these

¹²⁶ Famously, Dennett (1995, for instance) goes on to deny the mere conceivability of zombies: if this position is accepted, obviously Chalmers' and Kearns' & Magidor's arguments fail.

immaterial objects? Are they made of ether, of light? Are they spirits, do they have emotions? And, most importantly, *why shouldn't we argue that semantic facts supervene on the sum of physical and immaterial objects?* Let me stress this point clearly: if God had to add both physical facts and "spiritual" facts in the world, why don't we simply imagine that the semantic facts Kearns & Magidor assign to the ghosts supervene on all fundamental facts, *à la* Soames? In conclusion, the big difference between the Zombie World and the Haunted World is that the second is not metaphysically innocent, as it *adds* further objects to the world, thus multiplying the questions about it and the chance to explain meaning in new, maybe unfathomable, ways.

Though the idea of a Haunted World seems hard to swallow, then, let us assume for a second that it is possible for ghosts to exist. The second problem with this account is that, in a similar fashion to the recent dispositional attempts, it takes for granted that semantic facts exist in the first place. Obviously, a primitivist like Hattiangadi would be granted the benefit of having premised that we cannot do without meanings in order to even be able to write down intelligible concepts. But here, the argument is purely metaphysical: Kearns & Magidor are not directly answering KW's problem in their work, their primary goal is to make a semantic equivalence with Chalmers's proposal. Thus, assuming that semantic facts exist, after KW, is another not obviously true premise.

To be fair, Kearns & Magidor do briefly cite WRPL as a challenge to their theory (pp. 335-337): "Nothing in speakers' use facts is sufficient to determine whether by 'plus' they mean addition or quaddition". But this, as we have repeatedly seen, is just the tip of the iceberg: KW's conclusion is that talk of semantic facts is simply impossible to frame in a coherent way, hence they don't exist. Now, the metaphysical threat posed to meanings by KW gives a lot more trouble to the possibility of Haunted World than Kearns & Magidor are willing to consider, this time not because of the implausible existence of immaterial beings, but because of the implausible existence of *semantic facts* themselves. Once again, Chalmers's claim is tenable because of this crucial difference: we cannot question the existence of first-person phenomenal facts and we have no good reason to doubt the third-person phenomenal facts; on the contrary, we have all the reasons we need to ask semantic facts to buzz off from our ontology. The first argument for semantic sovereignty is, I think, refuted.

Let us turn to the second argument in favor of semantic sovereignty. Here is the outline: take two physically identical possible worlds, W1 and W2. Say they contain a girl, Jill, and a boy, James. Since the two worlds are physically identical, whenever Jill calls for James, she does so in both worlds. But, and here's the kicker, Jill in W1 will *refer* to the James living in W1, whereas Jill in W2 will refer to the James living in W2. Thus, semantic facts are different, physical facts being equal. It seems to me that the core of the problem here is found in the notion of haecceity as applied to "possible worlds". In a nutshell, the idea of

“possible worlds”, unless modal realism is accepted¹²⁷, should be intended as a *metaphor*: a more correct way to express it is to talk of “possible alternative states of affairs”. To put it differently, possible worlds are counterfactuals: Chalmers’s zombie world is not somewhere, but, according to him, it could’ve been (in place of the real world). Thus, W1 and W2 are not two separate worlds, even under an haecceity point of view: they are two alternative states of affairs, where the same fact happens, namely, that Jill calls for James. Clearly, then, the referent of the word “James” is one and only one: James. Additionally, the reasoning doesn’t work even if we see James as the same person, but with alternative characteristics: the utterance “James” still refers to the same object/person, notwithstanding their properties in the state of affairs under scrutiny.

Because of this, I think the haecceity argument has no power: if there is anything like meaning-I (or like meaning-S, for that matter), it is one and the same for the same physical facts.

It is now time to look into Hattiangadi’s (2019) attempt. Her introductory notes show more familiarity with KW’s claim, which is not surprising since she is among the most important scholars on the topic. Thus, she explicitly recognizes that KW did not *just* show that semantic facts don’t supervene on physical facts, but also that they don’t exist at all. That said, her position is that, while the first part is correct, KW’s conclusion is false: “There exist semantic properties that do not supervene on anything else” (p. 2). In Hattiangadi, the *epistemic fundamentality* of meaning is also more explicit: she outright argues that “semantic nihilism [what I call “meaning indeterminism”] is *prima facie* self-defeating: if no sentence has a meaning, and if the truth of a sentence depends on what it means and how the world is in relevant respects, then no sentence is true, including the sentences that make up the premises and conclusion of the argument for nihilism”. For simple as it sounds, I find that this is, in general, the strongest argument against meaning indeterminism in the literature to date. It does look like, as I, a meaning indeterminist, write down my words, I expect them to be followed by the reader in a *precise* way. I make claims, reject theories, evaluate arguments. But how can any of this be firmly rooted in any way? For now, let us just say that, though I ultimately disagree with Hattiangadi and I will defend against her accusations in Section Four, her attack is perfectly on point, unlike many others, and brings to the surface an *urgent* problem meaning indeterminists need to address convincingly if they want their theory to survive.

Hattiangadi also challenges the, seemingly straightforward, idea that primitivists cannot account for the nature of the semantic facts, as they are primitive. She takes it that the primitivist can actually account for both their metaphysical nature *and* our epistemic access to them. This also helps us understand that Hattiangadi mainly has in mind the

¹²⁷ Modal realism is the claim, notoriously defended by David Lewis in his (1986), that possible worlds actually exist. The thesis is widely rejected by the philosophical community, since it seems to most implausible and incoherent, due to unduly overlapping with the concept of multiverse.

philosophy of mind question rather than the metasemantic one: “if [the speaker] means addition by ‘plus’, a finite state of her mind stands in a relation to the infinite extension of the addition function”. On the epistemological side, she argues that suspecting the speaker may be in a deviant scenario (such as *Chair-2030*) is a skeptical sort of fallacy: if there is no reason to think that, we can assume the speaker is in the “normal” scenario of meaning *Chair* by their word. It should be said that this position is compatible with primitivism, but it doesn’t really answer Kripke’s worries: claiming that meaning-I is a connection between a finite state of the mind and an infinite function does not *explain* how this connection is supposed to work or to even be there. Analogously, saying that the speaker is warranted to assume the normal meaning-I instead of a deviant scenario in no way solves the *Privilege* question: why is *Chair* supposed to be the normal scenario and *Chair-2030* the deviant one? Thus, our final word on the mysteriousness of primitivist meaning determinism is this: that the primitivists can indeed account for what the semantic facts they are talking about are, but it remains an insolvable mystery *how* meaning-I is supposed to work or to prefer one meaning-S to any other. In conclusion, though Hattiangadi gives a plausible reason to think that a primitivist can explain meaning-I in different terms than itself, she does not give us a justification to think it is supposed to exist in the first place, other than by a supposed impossibility to satisfyingly account for communication otherwise.

Naturalized primitivism

When seen alongside dispositionalism, it becomes apparent that primitivism is of a very different breed. Of course, they are both attempts to show that meanings exist, but the way they approach the problem is almost specular. Starting from the same premise, “Meaning exists”, they take two opposite roads: the dispositionalist believes there is something in the world by which to explain it, whereas the primitivist accepts the impossibility to explain meaning in terms of anything else, and because of that claims its fundamentality. Primitivism doesn’t try to show how a different fact in the world could explain meaning, but how there is no way to do without meaning and, since it *can’t* be reduced to anything else, that it must be taken as primitive. Various ways to do so have been tried over time, but, as we have seen, the most promising ones, and the most true to the notion of meaning determinist primitivism itself, are the arguments in defense of Semantic Sovereignty. I have shown, hopefully, that the great challenge we need to take on as meaning indeterminists is to show that the premise “we cannot do without meaning-I” is false. Without this postulate, as I mentioned in the first Section, primitivism loses its very purpose, given that the entire endeavor begins from the recognition that meaning is unexplainable but also inevitable for us to do something constructive when we speak. If only we can give a plausible reason to think that communication -as we know it- can happen without a precise function for every word, that specifies the correctness of use for any possible instance of use, then we have defeated the primitivist challenge. Fortunately,

it turns out that we can, and that is all it takes to reject primitivism and accept the meaning indeterminist position, assuming that no other meaning determinist account (such as the dispositionalist one) will ever come up with a proper explanation of meaning, without the notorious contradictions and inconsistencies of the current hypotheses.

I have previously claimed there is something true about both dispositionalism and primitivism, that the two views grasp something about communication, though not the crucial aspect of meaning-I or of meaning-S. This is *a priori* plausible for two reasons. First, due to the Principle of Humility, it seems absurd to imagine that professionals who studied KW's riddle for years, if not for decades, should *completely* miss the point of its challenge. Secondly, we have way too many intuitions about how communication works that imply the existence of meaning, thus, although not in direct terms, they must be accounted for. My prime goal in this essay is to find space for these answers to KW in a meaning indeterminist account based on evolutionary considerations. I have already explained dispositionalism in terms of desire to communicate; it is now time to see what should be saved from primitivism.

In primitivism's case, I believe the silver lining of the theory consists in pointing out that we use our words as if we were¹²⁸ realists about meaning, both -I and -S. Primitivists like Hattiangadi focus on the impossibility of fleeing the semantic realm: whatever attempt we make, we always come back to the beginning, that is, the necessity of meaning. But it is a staple of contemporary philosophy of language and metaphysics, and for good reason, that, by definition, facts do not depend on what we think or say about them. This seems to be enough, in my eyes, to settle that there are some truths, regardless of whether or not we are able to express them linguistically. Thus, meaning indeterminism may be metaphysically true, even though, according to the primitivist, it is linguistically impossible to outline, due to its paradoxical nature. I am going to assume that this is true for the purposes of this paragraph, but I will also try to show that our being radically dependent on some "language game" that forces us to consider words as carriers of meanings-S does not make that true in metaphysical terms.

Recall what Miller (2020) says about semantic quasi-realism: it is an attitude we hold towards meaning, such that we make-believe that it exists even though it doesn't. His idea is that KW's approach to sentences such as "John means *Chair* when he says 'chair'" is analogous to Blackburn's approach to sentences such as "It is wrong to kill"¹²⁹. As Miller explains it, Blackburn's quasi-realist claim "is that we should not begin our attempt to explain the nature of moral judgement [...] by isolating states of affairs that we deem to be distinctively moral. Rather, we should [...] think of moral judgements as expressing states

¹²⁸ I am of course assuming that most people don't belong to any philosophical party about the nature of meaning, but they still use words in a certain way.

¹²⁹ Though it is arguably irrelevant to my claims about the effectiveness of quasi-realism in the realm of metasemantics, I want to make it explicit that I do not support Blackburn's quasi-realism about morals.

of mind whose function is not to represent anything about the world [...] and restrict ourselves to a methodologically naturalist toolkit”¹³⁰.

Miller claims this is a correct reading of KW’s thesis, but, regardless, we can definitely find it to work for our model of meaning indeterminism and to make sense of the “primitive” aspect of it. Let us apply Blackburn’s metaethical project to metalinguistic sentences. Firstly, *contra* meaning determinists, the point of the analysis is not to “isolate states of affairs that we deem to be distinctively [semantic]”. As I have repeatedly pointed out, in fact, assuming the existence of semantic facts makes the conversation reach a dead-end, because the point of contention is being dissipated from the start: one cannot discuss with a meaning determinist about their position on meaning, when the meaning determinist postulates meaning facts to make their case. Thus, a genuine discussion on the nature and existence of meanings requires a neutral methodology: we need to look at the facts we agree exist (namely, physical and phenomenal facts, or anyway facts that are not the point of contention) and look for a ground to semantic ones *as a conclusion*. With this in mind, we can point out another plausible candidate for what constitutes our “meaning talk”, alongside the desire to trigger behaviors, as seen in Section Two.

What we actually need to take as “primitive” in communication is our *psychological attitude* towards sounds and written text, i.e., it can be explained in terms of something else (in this case, evolution) but it can’t be logically *justified* in any way. This is the fundamental gap that doesn’t allow the “classic” answers to KW to succeed, KW’s solution itself included: they all look for a logically stringent answer that doesn’t leave room for the possibility of things working out differently, because they imagine meaning as something to be defined in terms that are detached from the context it sprouts in. Now, there is a good reason why many philosophers have run aground here. These psychological attitudes are the grounds for language games of the form: if I say *x*, you do *y*. Since, usually, sender and receiver communicate successfully, over time language games of this sort have shifted in our minds to become something similar to rules written in stone: meanings-S. But, as it may be apparent, there is no need for such things to really exist: it has become a part of our language games to believe language games to have ontological substance, but this is a mere fiction, since we cannot get our heads around it. As an equivalent, compare this with fictional stories or characters: as Kendall Walton brilliantly stresses in his (1990), fictions are not possible objects/worlds: they are imaginative prescriptions and, as such, they *can also* be inconsistent or impossible (p. 90). This is important, because if we accept the idea that meanings are fictional in nature (we make-believe that they exist), there is no need to make strictly logical sense of them anymore.

In recent years, a number of evolution theory scholars have approached the birth of meaning in *teleosemantic* terms: the meaning of a term is to be found in the reasons why it

¹³⁰ Miller (2020), p. 14.

evolved in such and such way, or in the reasons why it was maintained in a society (as suggested by Godfrey-Smith (2012)). Brian Skyrms and Jeffrey Barrett address this line of research in their (2019), arguing that meaning should be understood as “content that has become *ritualized* and *decoupled* from the relevant contexts” (p. 37). Once again, we ought to take this ritualization in entirely naturalized terms: as all animals do, we try to coordinate with others; as the most cognitively advanced animals, we do that with a potentially endless combination of sounds and symbols, which we naturally tend to associate to the same things over and over again. Hence, the legend of meanings-S and of our power to attach them to our utterances (meaning-I), is born.

Skyrms & Barrett’s simple, yet very effective claim for our current purposes, is that: “There is a practice of signaling with information transfer that settles into some sort of a pattern, and eventually what we call meaning or propositional content crystallizes out” (p. 34). It seems intuitive, after all, to imagine that when a sound has reached our interlocutor in the desired way, such as a scream to warn of some danger, it will be used again in future instances where the sender needs to communicate with the receiver. This is the process Skyrms & Barrett call “crystallization”: the existence of fixed meanings is generally a white lie, because it is optimal for fast coordination purposes. Let us see how this process works¹³¹.

Before the crystallization is in place, sender and receiver cannot be properly said to be in a “*language game*”. Rather, they start playing “a simple 2x2x2 Lewis signaling game”: here, we find two possible states of affairs (A: *predator*/ B: *no predator*), two possible signals sent by the first speaker (A: *scream*/ B: *no scream*), and two possible responses by the second speaker (A: *flee*/ B: *no flee*). The “game”, in a nutshell, consists in coordinating according to the speakers’ interests, which we will assume to be, in normal circumstances, to flee when the predator is present and to stay at ease when there is no threat. Also assuming that the first speaker normally wants to be sincere, its interest will be to communicate the correct information. Eventually, state A (*predator*) will be associated to signal A (*scream*), which will produce response A (*flee*); state B (*no predator*) will instead be associated to signal B (*silence*), which will produce response B (*ease*).

This result is “guaranteed with probability one”, as long as the community of speakers survives. The mechanism to get there is quite simple: upon seeing, for instance, a predator, the sender will choose one of two families of signals at random; equally random will be the receiver’s response. Every time a signaling system has been successful (has gotten the speakers to a desirable outcome), each element of the coordination chain will be reinforced; otherwise, they will simply start over again from scratch. When both speaker and receiver are consistent in coordinating successfully, the “basic game” can be said to be complete and it is the point where the crystallization happens. This is because the

¹³¹ All quotations are from Skyrms & Barrett (2019) if not stated otherwise.

coordination is *so* successful that both sender and receiver implicitly convince themselves (or are explicitly convinced by their community) that the signal holds a meaning in the strong, meaning determinist sense. Again, this result is most likely functionally optimal: if the speakers believe that a scream holds, in some sense, the concept of danger and the exhortation to flee, this will make the receivers' response faster and more consistent, so much so that the community enters further levels of interplay in communication, such as the language game of lies. In this case, the imposters will use the crystallized signals because they know with reasonable certainty that they will be followed by a specific behavior: if the deceiving monkey screams, the other members of the group will run away and it will be able to steal food, as some capuchin monkeys do¹³².

I think we ought to draw two important conclusions from these studies. The first one is the confirmation that we can give an evolutionary account of quasi-realism about meaning, which also seems to be more consistent and less problematic than its competitors; the second is that there is a reason why human communication is intuitively taken to be grounded on meaning-I by its partakers: since the layers of complexity in our communicative practices go much deeper than the basic game (or, for that matter, even than the first-level deception game), we tend to see our words as beacons for exact functions that have nothing to do with what really happens in communication. The origin of language is so deeply buried under our communicative subtleties that we needed a lot of philosophical work and scientific research to understand its nature.

Thus, after undergoing the naturalistic turn, the sort of primitivism Kearns, Magidor and Hattiangadi support should be turned into the claim that we, as speakers within a language game, may find it hard, or even impossible¹³³, to use the language game of meanings to argue that meanings are not there, but that it is metaphysically and conceptually true nonetheless. Remember, in fact, that a major worry for the primitivist was that it seems impossible to use words to say that words have no meaning, because of the internal contradiction that implies. That turns out to be a half truth: the crystallization of usage in our minds makes it *very hard* to abstract from it and to jump from the realm of pragmatic semantics to the realm of metaphysics, where meaning is nowhere to be found. On a negative note, though, once we have accessed a different perspective on what meanings come down to and how they function, it becomes very possible to jump outside the crystallization and to critically discuss the process in scientific and philosophical

¹³² As documented by Wheeler (2009).

¹³³ Even though I leave the benefit of the doubt, I am quite confident that we can express this concept coherently within the language game. After all, the very words I am using now seem to make sense, and they are conveying that concept; then again, the two ideas can be accommodated by arguing that, maybe, the way words are employed after being convinced by the meaning indeterminist position may be different on a psychological level.

What I certainly take to be true from naturalized primitivism is that we usually prefer to see meanings as "objects" in a strong sense and that this is optimal for "ready-made" communicative practices.

terms. This should also explain why meaning indeterminism is *prima facie* so counterintuitive: from within the language game we have a hard time understanding a concept we have been “blind” to in our linguistic practices for our entire lives; but as soon as the new framework is adopted, the language game’s nature becomes crystal clear, whereas it remains shrouded in mystery if we give in to it, as the primitivists do.

In conclusion, I claim that primitivism is a useful theory if it is understood as a quasi-realist claim about meaning-I and meaning-S. It is useful as a philosophical claim because it provides us with a piece of the puzzle that a mere meaning indeterminist perspective might lose along the way: the fact that, as speakers, we may not be able to get rid of the myth of meaning, and, importantly, that such myth is useful for effective communication and should not be discarded. Such approach would also clarify much of the differences among philosophers who walk a fine line between primitivism and meaning indeterminism, precisely because they may be more tempted by the supposed epistemological impossibility to get out of the vicious circle of meaning-talk (the primitivists) or by the inconsistency of meaning-talk (the meaning indeterminist perspective): it seems a good deal to let both theories (alongside dispositionalism, in the terms we outlined earlier) coexist in a coherent way, instead of having to choose to throw away one of them in its entirety.

Conclusions

We have reviewed an account of meaning that, though less popular than dispositionalism, has some followers in the literature and, in my eyes, has some *prima facie* strong philosophical appeal. Such appeal can be found in the ability primitivism has of accounting for both the intuition we have about the existence of meaning, and for the strong arguments KW, and the meaning indeterminists after him, have put forth against its analyzability. The primitivist stance is able to do this, but at a high cost: it asks us to accept that we cannot analyze meaning in terms of anything else, or, at the very least, that we cannot *explain* how and why it works; in order to preserve the first intuition as well, it also asks us to believe *a priori* that meaning-I is there, without further warranty other than its seeming irreplaceability.

We have then pointed out that the main misunderstanding behind this theory is the idea that a framework of communication that can’t account for the existence of meaning-I is in some way flawed, or less appealing than one that does. To be sure, it seems intuitively impossible for the meaning indeterminists to give us a credible reason to believe that what they argue for is true, when what they argue for is *meaningless*. This argument turns out to be dismissible when we notice that one such theory, the meaning indeterminist one, doesn’t account for meaning-I simply because it doesn’t exist: what we are trying to do is to give an account of communication, where a linguistic content *is* sent from a sender and received by a receiver, but where the notion of meaning-I, in its strong sense, is not

essential to the endeavor anymore. Something similar has become recently popular in the context of perception: direct realism is a tenable position, and an attractive one, precisely because it makes without the problematic notion of sense data; there is no good reason to not do the same here.

Far from being a flaw in the theory, then, the missing piece of meaning-I is just a more correct description of how the world is, rather than a weight the meaning indeterminists must bear if they want to support it: getting rid of the middle step makes the account of communication sharper, less convoluted. In fact, not only is the meaning indeterminist theory giving a (post-KW) plausibly more correct account of “what there is”, but it is also offering, as we are about to see, an alternative account where humans (and, importantly, other creatures) communicate, thus saving an intuition, that of meaningful communication, that Kripke himself seems very concerned¹³⁴ about losing.

We have also evaluated two attempts to systematically defend semantic primitivism: Kearns & Magidor (2012) seem to fail on both lines of argument (the Haunted World and the haecceity argument); Hattiangadi, instead, proved a more insidious opponent. In a nutshell, the reason why Kearns & Magidor seem to fail in their attempt is because they profusely use the notion of possible worlds, without accounting for a series of difficulties: it is not clear whether it is really *possible* for immaterial beings to exist *and* mean something by their “words”; it is not clear why it would be impossible to argue that such immaterial entities, plus the material world, are not enough to claim a supervenience relation between semantic facts and the Haunted World; it is not clear why haecceity should matter in a mental experiment such as the “possible worlds” one. On the contrary, Hattiangadi’s strength has been her less systematic metaphysical account, and her more stringent focus on epistemological problems that arise from the acceptance of meaning indeterminism: in short, she finds that a meaning indeterminist claim, if accepted, cannot, by definition, have a precise meaning, thus pointing out the viability of the primitivist claim is more than enough to take it seriously.

Hattiangadi’s fundamental line of attack against meaning indeterminism has been that its supporters have no way to grant that their words mean what they mean, and thus that the truth value of their arguments is completely up in the air. Notice that this line of attack is slightly different from the one we have started from: she is not necessarily saying that meaning-I is essential to communication, but, less strongly, that the exact conveyance of a linguistic content (a proposition) is not possible within the meaning indeterminist framework: since meaning indeterminism accepts the proposition “meaning-I does not exist”, such proposition needs to have a *very precise* meaning, which it cannot have.

Because of all this, she found it more reasonable to postulate meaning-I as fundamental, and outlined a series of precautions to make the, still radical, claim more digestible.

¹³⁴ This is what he implicitly gives off when he calls KW’s proposal “crazy”.

Among them, we have found an interesting line of defense that makes meaning-I at least intelligible (her primitivist claim thus turns out to be more metaphysical than epistemological), a claim of Chalmers-like flavor about the conceivability/possibility of semantic zombies, and the idea that it is a skeptical fallacy to suspect being in the deviant semantic scenario. I have addressed these propositions, not always agreeing with them.

What I haven't given a proper reply to, is Hattiangadi's main line of offense; rather, I admitted that her worries about meaning indeterminism are also my worries about meaning indeterminism. It has, indeed, always struck me (and many philosophers before me) as highly problematic to not be able to assess exactly what people mean by their words, when accepting meaning indeterminism. But I now believe that these worries can be silenced by giving a complex account of how communication works, inclusive of a reason to expect that any English-speaking reader should understand what I want to say by my words, or, as we usually say, what I *mean*. The keystone to solve the riddle is Brian Skyrms' "salient equilibrium", which we will look at, among other ideas that give the meaning indeterminist account more solidity, in the following Section.

4: KW, naturalized

We have been looking into a wide variety of meaning determinist responses to KW's fundamental problem of meaning, but none of them has been found to be satisfying. In this section we will present and assess the two main meaning indeterminist accounts, KW's own communitarian answer and the evolutionary answer, concluding that the second one is a full and sufficient philosophical disentanglement of our Paradox.

The "naturalized" solution I will support in this section agrees with KW on two fundamental points:

- (1) That there is no such thing as meaning-I (i.e., meaning indeterminism is true);
- (2) That none of the meaning determinist answers discussed in WRPL are proper answers to the challenge.

Additionally, though we have seen that Kripke is not clear or explicit about it, I also endorse KW's claim about meaning-S, such that:

- (3) The potential existence of meanings-S is irrelevant to our communication practices.

On the other hand, there are two important points where my position diverges from KW's theory:

- (1) That it is not true that "Nothing about 'grasping concepts' guarantees that [communication] will not break down tomorrow"¹³⁵;
- (2) That it is not true that "[Speaker A] will judge [speaker B] to mean [x by y] only if they judge [B's] answer to [...] agree with those *they* are inclined to give"¹³⁶.

Let us start by giving proper definitions to the concepts we will be dealing with.

Communication is coordination

I have been arguing¹³⁷ over the course of the essay that it is possible to make sense of communication without ever resorting to the "strong" notions of meaning-I and of meaning-S. It is now time to try and justify more fully this idea, by suggesting how this could be done. The fundamental idea we will draw fully from can be found in the works on game theory by David Lewis (1969) and, more recently and with a more empirically grounded approach, Brian Skyrms (1996, 2004, 2010): it is the idea of *coordination* among individuals, specifically as it emerges after the workings of natural selection. We constantly find ourselves in the presence of coordination problems in our environment: these can be understood as states of affairs where "Two or more agents must each choose one of several alternative actions. [...] The outcomes the agents want to produce or

¹³⁵ WRPL, p. 97.

¹³⁶ WRPL, p. 91.

¹³⁷ Following Guardo's (2019).

prevent are determined jointly by the actions of all the agents. So, the outcome of any action an agent might choose depends on the actions of the other agents. That is why [...] each must choose what to do according to his expectations about what the others will do”¹³⁸. We live in an environment where we want to (have a disposition to) have some facts happen; sometimes we want to act ourselves with the help of others, some other times we want others to act in a specific way. The second of these dispositions is most successfully realized via communication, which in human practices usually is tantamount to using a language.

I have mentioned that the notion of “salient equilibria” is the key to success for a naturalized account of meaning and communication to work. An “equilibrium”, as defined by Lewis (1969), is a “combination in which each agent has done as well as he can, given the actions of the other agents” (p. 8). An equilibrium, then, is not necessarily an optimal turn of events for any of the participants in the game: it is the optimal turn of events each “player” can produce, given the actions of the others. In order to dissipate possible misunderstandings, let us see a few examples.

Say Paul and Mary have a coordination problem: they want to meet for a coffee, and this desire has a +1 value for each, if satisfied. Assume, in addition, that they usually meet at three different coffee shops. Clearly, there are three possible coordination equilibria in this scenario: the three states of affairs where both Paul and Mary go to the same coffee shop, thus meeting each other and meeting the requirement for solving the coordination problem. Without further specifications, we should assume that all equilibria are equal: as long as they meet at a coffee shop, their happiness – which we can schematically visualize in terms of the +1 value - will increase in the same way, no matter the scenario. So, let us complicate our picture slightly. This time, Paul and Mary want to meet again, but they also have preferences concerning the location: they both would rather go to the coffee shop that is nearest to where they live. Assume the bars are equally distant from both their houses: in this case we will have a value distribution of the following kind.

- They don't meet; Value 0
- They meet at the closest coffee shop; Value +2
- They meet at the next closest coffee shop; Value +1,5
- They meet at the farthest coffee shop; Value +1

This time, we still have three equilibria, but they are not of equal value, that is, we can objectively pinpoint the optimal one, the inferior one, and the one in between. Obviously, since these are very simple scenarios, they are usually very easily resolved in communication: since *both* players have the same preference, they will most likely solve the coordination problem by saying “Should we go to the nearest coffee shop?”. But, theoretically speaking, the third and fourth scenarios are still examples of equilibria; no-

¹³⁸ Lewis (1969), p. 8.

one got their favorite outcome, but neither player could have acted more optimally if taken by themselves.

Finally, let us complicate the scenario a bit more still. Paul and Mary still have the same objective in mind, but their preferences on coffee shops diverge, because the one closest to Mary is farthest from Paul. In this scenario, there is no optimal solution to the coordination problem, but there are optimal solutions from the perspectives of each player. There are still three equilibria, but two of them are competing ones.

Now, let us turn to what makes an equilibrium *salient*. Again, we start from Lewis's description of the concept: "[a salient equilibrium] stands out from the rest by its uniqueness in some conspicuous respect" (p. 35). Importantly, salience should not be associated with pleasantness: what stands out can do so for a number of reasons, such as its size, its commonness, even the fact that it is considered unusually *unpleasant* by the speakers. Say that Paul and Mary are at an exhibition of paintings. Among others, they see a painting they deem quite beautiful and one they despise so much they keep laughing at it for some time. Say the next day Paul texts¹³⁹ Mary: "Remember that painting from yesterday?". It is likely that, given the *impression* the hideous painting left on them, Mary will immediately know (or, at least, suspect) that Paul is referring to *that* painting and no other, even though, in a meaning determinist flavor, the meaning of his words applies to any of the paintings they saw at the exhibition. Whatever the reason, what must be remembered is that salience is an external, psychological factor that makes one of the equilibria especially likely for players that do manage to coordinate.

Having introduced the focal point of our solution, it is now time to make two important *caveats* here, in order to defend this argument on solid grounds the reader may or may not already agree with. I am hereby assuming that naturalism is the correct metaphilosophical view of the position philosophical reflection has in the world¹⁴⁰. I don't see good arguments to reject this idea, or arguments that don't fall into a very unwarranted skeptical or subjective-idealist vein, but still, rejecting this premise makes the "Semantic Naturalism" account hard to accept solely on the bases I am providing. In other terms, I am taking for granted the fact that humans are animals that, due to evolutionary and social circumstances, developed a series of intellectual faculties and practices, including language and philosophical enquiry. If this is not taken to be true, then the only good reason the denier may have to support my claim is that no other is consistent enough on its own terms.

¹³⁹ This is important to prevent non-linguistic factors from intervening in the example.

¹⁴⁰ Not that this should be automatically taken as an indication of its correctness, but it ought to be noted that naturalism is also the most popular option as a metaphilosophical framework among professional philosophers (Naturalism: 49,8%; Non-naturalism 25,9%; Other 24,3%). Source: Bourget & Chalmers (2014).

The second assumption I will make is more modest, but essentially dependent on the first: that human language stemmed from coordination practices in primitive hominine communities, following Tomasello's (2014) study, among others. Again, I believe this theory to be solid, but the matter cannot be addressed in full here, and it is not quite uncontroversial, though it seems to be currently taken as an obviously true claim in some evolutionary sciences circles¹⁴¹. With that out of the way, we can take a closer look at the essential features of coordination in a given environment, and how they work together.

There seem to be two families of approaches for an individual to try and coordinate successfully with others. The first, less effective but still very common one, would be to simply predict, based on previous experience, inferential thought and/or sheer luck, what others will do, and behave accordingly. For instance, a wolf may expect the wolfpack to attack a prey from the sides and decide to attack it from the rear, in order to make the coordinated attack more deadly. But when decisions are made on a less instinctual basis, especially in human environments, these predictions can turn out to be not consistent enough. Thus, the second, more optimal, approach to solving coordination problems involves the employment of "signaling systems", such as cries, gestures, drawings, words, which send a message to the receiver before the action takes place. Signaling systems are a powerful solution to coordination problems because "the sender has private information that she wants to send to [...] the receiver"¹⁴² and a signal both creatures share as meaningful makes the transition possible. The reason why I call this a more optimal approach should be clear: if we agree on a conventional sound/sign/word to represent a fact x , the coordination problem is solved much faster and more efficiently, compared to a situation where we always start from scratch trying to work out a situation together. Even in the wolfpack context, if the wolf attacking from the rear used a specific howl to inform the pack of its action, the coordinated attack would most likely be better organized. As this example should emphasize, the agreement need not be explicit: "Explicit agreement is an especially good and common means to coordination, [but] we do without agreement by choice if we find ourselves already satisfied with the content and strength of our mutual expectations" (Lewis 1969, p. 35).

Lewis also finds "higher-order expectations" to be a fundamental aspect of coordination, a piece of the puzzle that cannot not be there for the event to happen consistently. But, though they definitely *improve* the chances of coordinating successfully, it is important to make it clear that they are not *necessary* for normal coordination and, thus, communication, to take place. Actually, it is much more plausible that in the overwhelming majority of cases there is no higher-order expectation at work at all: as in

¹⁴¹ See, for instance, Skyrms & Barrett (2019): "We all think that humans are animals, that human language is a sophisticated form of animal signaling, and that it arises spontaneously due to natural processes".

¹⁴² Skyrms (2014), p. 82.

Guardo's (2019) "Qlock" example¹⁴³, worker Alpha only needs to *desire* a block and to be psychologically disposed to call it "Qlock" for him to communicate in this way to Beta; analogously, Beta only needs to desire helping Alpha and to be psychologically disposed to point his attention to blocks whenever Alpha says "Qlock" for their communication to be successful.

Similarly, it is important to discern my use of the notion of "natural" communication practices from, as an eminent example, Grice's (1957). In his famous article, "Meaning", the author makes a distinction between natural and non-natural meaning. In his terminology, the natural use of words or other signs is such that if x means that p , this implies the truth of p : they are a *direct* indication of how things are in the world. A non-natural use of signs is such that if x means p , this does not automatically imply the truth of p : the sentence "the snow is white" means that the snow is white, i.e., it directs our attention to a certain proposition, but it does not automatically imply a specific truth-value to be attached to it, not even an indeterminate one. An important trait of non-natural meaning is, according to Grice, the intention of the speaker to produce a certain effect: while natural meaning is simply associated to a fact (such as, by Grice's example, giving a man a photograph of his wife cheating on him), non-natural meaning is associated to a fact through the association to the sender's intentions (such as giving a man a *drawing* of his wife cheating on him). Clearly, in the first case the indication that the wife is cheating is primarily given by the *evidence* provided by the photograph: assuming the photograph is a *true* photograph, it is true that she cheated; in the second case, the indication is not given by the symbol itself: a drawing can represent anything at all, true, false, or even impossible. What communicates that the wife is cheating, in the second case, is the supposed intention behind the drawing and behind the action of giving him the drawing, so it must be found in the drawer rather than in the drawing, in the signaler rather than in the signal taken by itself.

Grice's account is two-fold, as the natural/non-natural dichotomy suggests. On the one hand, he has in mind something similar to Lewis's higher-order expectations: we need to be aware (on a more or less conscious level) of the senders' intentions in order to get the non-natural meaning of what they say/write, because the content of their message cannot be automatically derived from the evidence alone, but it must be contextualized within the sender's communicative intentions. On the other hand, the point of his distinction seems to contrast with Lewis: while some signs are intrinsically connected to some facts, such as the dots that imply measles (so that we are warranted in saying that "These spots *mean* measles"), others are just conventions, which can only be appreciated in their meaning if we are able to "read the speaker's mind", so to speak. I don't believe this clean cut to be entirely warranted, and I mean this as a generalized point about meaning and not as an exception-based argument, but, to be fair, Grice grants that "[not] all our uses of 'mean'

¹⁴³ Pp. 122-123.

fall easily, obviously, and tidily into one of the two groups I have distinguished; [yet] I think that in most cases we should be at least fairly strongly inclined to assimilate a use of ‘mean’ to one group rather than to the other”.

I think that, in his terms, Grice is right. The way I use the notion of “naturalness” in linguistic practices will be quite different, though. The way in which I will employ the word is based on the idea that *all* linguistic practices are naturally based: the expressive properties words have are, in a way, objective, that is, they have a *consistent* effect on the receiver, at least as long as the receiver shares a common biological nature¹⁴⁴ with the sender. Thus, I will not use the word “natural” in the way Grice does; instead, it will be a term aimed at showing that all of our linguistic endeavors are the result of an evolutionary process based in our animal nature. To make the difference clearer, take again the “cheating wife” example. Grice tells us that a (true) photograph of her cheating on her husband is a natural signal, whereas a painting of her cheating on her husband is not: that because the first signal implies the truth of its content, due to it being *directly caused* by the fact it portrays, while the second doesn’t. The way I will frame the naturalness of meaning is, instead, directed at the common nature of sender and receiver of the signal: in my terms, the painting of the cheating wife is still a natural signal, because the receiver (assuming he shares a similar enough nature and culture with the sender) *sees* in that signal the content the sender wants him to see. If their natures were too different, that would not happen, even in Grice’s natural examples: if I show a lion a photograph of his partner cheating on him, that would not be enough for the lion to coordinate with me on the information I am sending him.

Everybody does it

In his works concerning, among other aspects of evolution, the evolution of meaning, Brian Skyrms frequently stresses the fact that all animals communicate: “Birds do it” titles a section of his *Evolution of the Social Contract* (p. 86); even “Bacteria do it”, titles another in his *The Stag Hunt* (p. 55). In short, to use his words, “signaling systems are ubiquitous at all levels of biological organization”, though, of course, he does allow that “there is a great gap between human language and what we find in other animal species”. Let us see what, if anything, this is supposed to tell us about meaning-I and, possibly, meaning-S.

Notice that individuals are usually *strictly better off* coordinating than they are struggling by themselves to reach a goal: far from being just rhetoric, this claim is evolutionarily justified (creatures that live in groups tend to survive more and better) and is most likely the reason for the hegemony a weak animal such as the *homo sapiens sapiens* has attained over the natural world. Again, Tomasello (2014, pp. 170-175) has some strong evidence in favor of the primacy of cooperation in this endeavor: in particular, it seems very likely that

¹⁴⁴ It is not even necessary for this common nature to be particularly close: that is why, in some loose way, humans can communicate with dogs, cats and other animal species.

our massive tendency to teach others what we learned before and to pass on knowledge as a community has granted the extremely fast improvements human beings enjoy as opposed to all other living creatures. Naturally, this could not have been possible without a massive diffusion of more or less sophisticated communication practices and of devices for their widespread diffusion.

Unsurprisingly, then, at least a wide variety of animal species have some degree of communication: “birds use songs [...], bees have the dance language [...], monkeys have different kinds of alarm calls for different kinds of predators”¹⁴⁵. This may sound odd or surprising to some, but it shouldn’t. The ability to communicate with one’s peers is a predictable result of natural selection: if communication grants more effective coordination processes among members of a same group/species, groups and species that communicate successfully will, most likely, survive and prosper. What may be surprising is the fact that so many different species communicate, even evolutionarily primordial ones and with very limited cognitive functions, but the idea becomes more digestible as soon as we sketch more precisely what sort of communication that is. Importantly, we would call most of these behaviors *instinctual* or *psychological*: it would be giving too much credit to bees to think that they logically adhere to a rule whenever they “communicate the direction, distance, and quality of a food source” (*ibi*).

It seems to me that we can easily see this in animals such as insects because their kind of communication is intrinsically connected to the current state of affairs that needs communicating and, as such, it *does* lack something: it lacks what we properly call “language”. Crucially, as Di Martino (2017) underlines, “Language is not an accessory, it does not [only] intervene to express something that would still be already at hand without it” (p. 143, my translation). This is to say that a bee could still gain the information it needs from the world itself, were it not communicated by its peers: communication remains optimal because it speeds up the process or makes it possible for one bee to get an information it wouldn’t have received on its own, but the essential point is that the information that is being communicated is always *present* in the world.

Thus, what bees (and most other animals) completely lack is the possibility to “bring up” in communication concepts and objects that are not immediately related to the current circumstances: even monkeys, who are able to distinguish *which* predator is attacking them, are arguably not able to bring them up if such predators are not their immediate concern. Our ability to abstract from the current, actual state of affairs is both the result of our cognitive primacy *and* a cause of its further development (Tomasello 2014, p. 150). This is, most likely, the reason why human communication strikes us as entirely different from what happens in the rest of the natural world: most animals merely react to their kin’s behavior by behaving in such and such way themselves, and have other members of their

¹⁴⁵ Skeyrms (2014), p. 86.

species act in certain ways in return¹⁴⁶. This is especially clear when looking at *bacteria*: that is because, even though “[they] have developed and maintain effective signaling systems”, we can safely assume, due to obvious and hardly disputable cognitive shortcomings, that “they do this without benefit of *common knowledge, intentionality, or rational choice*”¹⁴⁷. Under this aspect, bacteria are nothing more than cogs in a machine “designed” (metaphorically speaking) by evolution: they happen to coordinate, but they certainly do not *know* they do, they are not able to bring up concepts and objects outside of their immediate environmental context and have a very limited amount of messages they can share.

The question whether other animals different from the human being can be considered, under this aspect, something more than that, is a hard one and it is probably not important for our study. It seems reasonable to me to concede that the communicative capacities of animals are more or less directly proportional to their cognitive abilities, and to assume that no animal other than the human being has developed such refined communication practices as to be able to own a proper “language” in the sense employed by Di Martino¹⁴⁸.

Let us now go back to our problem. I have sketched this naturalistic account in order to place human communication inside it, only to end up arguing that human communication does have something *more* than the other animals’ does. But accepting the, quite obvious, superiority of human language does not amount to saying that human beings should be thought of as having developed or as having connected to some *ontologically loaded objects*, whether they reside in our minds (meaning-I) or somewhere else (meaning-S). Rather, the naturalistic account is supposed to make it clear that, other than some refinements, impressive as they might be, we have *nothing* (i.e., no object, mental or otherwise) more than any other animal at the core of our communication practices: we have the same need (coordination) and the same solution (communication). The fact that we do it better, faster or more easily than any other living creature is just a matter of degrees.

The reader will have intuited what I am getting at: if the meaning determinists agree with the naturalistic background I have suggested, they owe a justification to explain why human communication is supposed to be *ontologically* different under this aspect. Granted that human language is far more complex and polished than all other animal

¹⁴⁶ One might want to argue that this is what I claimed about *human* communication at the end of Section 2. Notice, however, that I framed the disposition to behave and have others behave in such and such way as an important *component* of human communication (and, by now we can say, of communication in general), but I do not claim that it is the *only* feature of human communication or that it is what distinguishes it from other sorts of communication. Rather, I might argue with more conviction that it is the aspect human beings *share* with all other communicative creatures, or its most evolutionarily primitive trait.

¹⁴⁷ My emphasis.

¹⁴⁸ Another important and difficult problem here concerns the risk of “projecting” our more refined and self-conscious communicative practices on animals that may or may not have such subtleties in their communication (Di Martino 2017, pp. 24-26).

communication devices, and that it makes us capable to bring up objects and ideas that are not there or that may not even exist or be possible, the fundamental question is: why should we think this is anything but an evolutionary step in the line of animal communication? The only reasonable answer, to me, would be that something in the way we communicate cannot be addressed by the evolutionary studies on animal communication. In the next few pages, I will try to show that there is no such thing.

To sum up, I claim that there is no metaphysical difference between human communication and the other animals': in its essence, human and animal communication are exactly alike, their only difference being in the degree of complexity and nuance, given by our more developed cognitive functions. I will now address a few possible objections to the picture I hereby presented.

It is not obvious that the acceptance of this naturalistic framework should immediately have us abandon a meaning determinist account, even when recognizing all the scientific evidence available as true. First off, one might argue that if bacteria communicate, they do convey meanings-S, though without a meaning-I intentional act. We have seen that dispositionalism is a lot more powerful in the context of metasemantics than it is in the context of the philosophy of mind, because the epistemic asymmetry found between meaning-I and dispositions to use words in a certain way does not hold in the other case¹⁴⁹. Thus, the dispositionalist may have no problem to accommodate their theory to bacteria, too: the bacteria mean x in their communicative practices in ideal conditions (or other counterfactual, depending on the dispositional account) iff they follow the rule that consistently applies to x , whether they know it or not. The primitivists may find themselves in more of a plight here, since they usually tend to look at meaning-I, but they also could respond by arguing for the metaphysical primacy of meaning, *à la* Kearns & Magidor, or they could suggest that meaning-I is a mental state that is primitive but only a few creatures have access to it, namely human beings¹⁵⁰.

With all this, I still don't think dispositionalism or primitivism are tenable positions here. Aside from the familiar objections we have stressed in the previous chapters, there seems to be a further difficulty here: how does meaning-S have anything relevant to add to our picture, if it has no direct connection to what the bacteria do? Remember that meaning determinists strive to defend the existence of meanings because they deem them *essential* to communicative practices; instead, these arguments would merely find a way to let meanings survive, even though we clearly don't need them to be there for communication

¹⁴⁹ For a thorough defense of this point, see Guardo (2020a).

¹⁵⁰ After all, returning to the equivalence with the fundamentality of phenomenal facts, Chalmers does not need to claim that everything has phenomenal states in order to argue that phenomenal states may be primitive.

to happen. In other words: there may be a rule the bacteria follow to perfection¹⁵¹, but they most certainly don't *mean* it by their coordinative behavior. What is the point of assuming meanings-S are there at all then, by now? Ockham's razor looks sharp enough to do the job and get us to toss them away. The same, clearly, goes for meaning-I: why would we want to assume that only a select number of individuals who communicate have this primitive "power", instead of just flattening ourselves to the level of other communicative animals?

It seems to me that, as things stand, only one serious worry should be remaining before accepting meaning indeterminism, which stems from Hattiangadi's considerations. Recall that we took from Hattiangadi's (2019) to find it concerning that it seems impossible for a meaning indeterminist position to be able to communicate a *precise* concept to the interlocutor, without a corresponding precise meaning to our words. Let us compare this problem with the one we seemingly find answer to by looking at the other animals' communicative practices. Sure, bacteria, birds and monkeys do communicate with each other, but do they convey *propositions* in such interactions? It does not seem so. In fact, what we call "animal communication", as we have said, lacks the ability to bring up objects and concepts that are not present in the environment; but there is also something else: it lacks the *specificity* that words seem to provide. This is, to my eyes, the most powerful understanding of the primitivist objection: sure, animals may very well communicate, but we do something more; namely, we use one specific sound to suggest a specific rule to follow and not *just*, as other animals do, to cause a certain behavioral reaction.

Mind you, this is not exactly what Hattiangadi would say¹⁵², but it is what I take from her work as our priority to defend meaning indeterminism, since there doesn't seem to be an easy answer in the literature. My attempt to block this objection will employ the notion of triangulation and that of salient equilibria, as proposed, respectively, by Donald Davidson and Bryan Skyrms. Together, the two notions will grant us with a good way to account for the preciseness of human language.

Natural triangulation

To properly address the notion of triangulation, we need to return to the differences between Davidson's and KW's theories of meaning. It is important to stress that I am addressing the *later* works¹⁵³ by Davidson, which are, in fact, the centerpiece of his "triangulation argument": the reason for this is that I don't see his earlier works as proper

¹⁵¹ But, of course, if one is persuaded by KW, there are infinitely many ones compatible with the bacteria's behavior, taking us back to the starting point.

¹⁵² She sounds a lot more "extreme" in her argument, claiming that there is no way to give a truth value to our sentences in the meaning indeterminist account. I think I have given reason to discredit this strong claim. The problem about preciseness, though, may remain.

¹⁵³ Specifically, Davidson (1992), (2001a) and (2001b).

answers to KW's problem. I will briefly explain why I think that. In his earlier works, as he defends a first person authority of sorts concerning intentional acts, Davidson doesn't seem to have a correct understanding of KW's force or, at least, to not have taken it into account¹⁵⁴: when discussing the perspective of the speakers when they mean something by their words, Davidson says that "The presumption [that speakers cannot be wrong about what they mean by their words] is essential to the nature of interpretation - the process by which we understand the utterances of a speaker" (1984, p. 110). His contention here is that there is a fundamental difference between the sender of the signal and the receiver: the sender cannot question what he means, at least in an evidential sense, like the receiver can, because that would hinder the entire process of interpretation: in a somewhat primitivist flavor, the speakers' meaning-I is a presupposition for any interpretation of what they say. Importantly, the speaker can be wrong about what he means in this account, "but the possibility of error does not eliminate the asymmetry". Now, this "asymmetry" Davidson speaks of is, in a way, a prelude to his conviction that his account of triangulation can give a straight reply to KW. What the asymmetry consists of, in fact, is the fact that one needs to give the utterer's meaning for granted before assessing it, that is, that is all there is to work with. But, as we have seen, there is no need whatsoever to speak of *meanings* in this intuitive idea: what we must take as a given, subject to further interpretation, is the *behavior* disclosed by the speaker. Among the speaker's behavioral acts, there are also sounds which are usually associated to objects and/or concepts; but this is not to say that the meaning of the speaker's words should be taken for granted *a priori* by any party.

This account seems to fall prey to our usual problem: a meaning-I is immediately assumed to exist, and in this case it is thought of as a constitutive block for any proper communication (and potential correction) to happen. As I suggested, it seems that Davidson has a too stiff conception of meaning, even before such meaning has been properly assessed, and this is an approach he will persevere in: since we call a meaning indeterminist someone who denies the existence of a mental state of meaning something by one's words (in the philosophy of mind) and/or of precise rules of correctness "out there" (in metasemantics), Davidson's position will actually be compatible with it, but, as things stand in his earlier works, he cannot really help us defeat or accommodate KW's problem.

On the contrary, the late Davidson, especially in his early-21st century works, holds KW in high regard, nevertheless disagreeing with him. As I claimed in the first Section, it seems to me that their divergences are more formal than they are substantial. As Sultanesco & Verheggen helpfully state, when considering Davidson's rebuttal, we need to make the usual distinction between the question in the philosophy of mind and the one in

¹⁵⁴ Since WRPL had already been published at the time and Davidson does not address it in his (1984).

metasemantics¹⁵⁵. They also agree that the philosophy of mind question is hopeless in Davidson's eyes, which is also the conclusion KW comes to: there is no such things as meanings-I, because they cannot be made sense of and are inessential to a proper account of communication, thus "the expectation that the world contains such things as meanings, understood as entities accompanying uses, is incoherent, and can only be the result of confusion" (Sultanescu & Verheggen (2019), p. 26).

On the other hand, Davidson is glad to answer the metasemantics question positively, via the concept of triangulation, "the idea of two individuals interacting simultaneously with each other and the world they share" (p. 7). As Malpas (2019) argues, "although the idea may appear at first sight to be intended purely as a metaphor, the structure of triangulation seems actually to direct attention to the way in which knowledge, action, and content are fundamentally dependent on the genuinely embodied and located character of speakers and agents". In fact, one first essential aspect of triangulation consists in the causal role a distal object has on the focus of attention shared by the speakers¹⁵⁶: the reason I look at the moon when the astronomist points at it, instead of focusing on the scholar's index finger, is that both me and the astronomist are casually affected in a relevant way, given the context, *by the moon* itself, even before assigning it any associated symbol. The other crucial element is the interrelation between sender and receiver: it is not obvious that I would understand the meaning of "moon" in the same way if I wasn't in a specific relation with the astronomist that points at it; in general, a meaningful utterance is possible only when incarnated in a relational circumstance.

At this point, it must be said that Davidson is probably closer to this essay's position than KW is: the account I am supporting is, in fact, a meaning indeterminism that safeguards the (arguably indisputable) fact that there *must* be a reason why communication works for our purposes most of the time. Conversely, KW seems to conclude from the non-existence of meaning-I that "our license to say that we mean addition by plus is part of a 'language game' that sustains itself only because of the brute fact that we generally agree (Nothing [...] guarantees that it will not break down tomorrow)". I take it that with "brute fact" KW implies that the event of coordination when we talk or write something down is unjustifiable: not only do we not know how that happens, but there is no positive reason for it, which is why it may very well "break down tomorrow".

This is, in my view, KW's great mistake, and the reason why so many philosophers find his argument impossible to agree with. In this sense, I agree with the likes of Hattiangadi: it is both unrealistic and incoherent to try to *argue* for something, in written text, expecting

¹⁵⁵ I refer to Sultanescu & Verheggen (2019), p. 49-50. I take it that when they speak of "conditions of correctness to govern the use of expressions" they refer to the metasemantics question and when they speak of "objects in the mind that the agent may consult prior to her using expressions" they refer to the philosophy of mind question.

¹⁵⁶ See Davidson (2001b).

for it to be understandable in completely different terms than KW predicted: if what he says were true, then he should not be surprised if a reviewer takes his argument to be about the reproductive habits of butterflies, or whatever else. Saying that our communication is basically the product of pure chance and that it could in principle stop working tomorrow is just as bewildering as Hume's claim that our inferential way to determine the laws of physics is in principle mistaken because they *might* change tomorrow¹⁵⁷. Both claims, in fact, seem to confuse the notions of *possibility* and *probability*: how absurdly unlikely would it be for the world to have happened to work *as if* there was a causal relation between facts? And how absurdly unlikely would it have been for communication to constantly help us' coordinating *by pure luck*? It seems that the first good reason to doubt KW and Hume are right in their respective *skeptical* claims is that the amount of consistency in these events (successful communication and inductive reasoning) is so enormous that it is far more plausible for there to be a solid ground than for it to be a product of sheer chance. Clearly, KW is meant as a challenger to our presuppositions about meaning, and fortunately he is just a fictional philosopher. But, regardless, it is important to understand how philosophically unviable his solution is: it is actually plausible that, given his solution, his problem has been long perceived as more "threatening" than it really is.

Thus, Davidson's notion of triangulation helps out in this context. He frames, in a geometrically intuitive way, the ingredients we need to be able to communicate without fixated meanings: something that catches our attention, and someone whose attention is caught by that same thing. This is precisely the solid ground we need to argue that communication exists, because the word we begin to use in our coordinative practices has a referential power justified by natural causes. In the same way that we can argue that it is much more reasonable to explain inferential knowledge through the interaction of laws of nature, we can more reasonably explain communication through our similar psychological reactions to the world. At the same time, though, this is perfectly compatible with a meaning indeterminist position and is pretty much a different way to state the psychological condition argument: all it takes to communicate is to have a similar psychological approach to words and to objects in the world¹⁵⁸. Due to that, Davidson's worries on the matter seem unwarranted and the mere product of a terminological clash: after all, he agrees with KW about the non-existence of meaning-I, but he opposes the idea that we should not speak of meanings, since "for words to mean what they do [...] it is for them to have been used by speakers. [...] To be a speaker is to have been understood by another in triangular interactions that involve not only the shared world, but also its recognition, by both triangulating creatures, as a world"¹⁵⁹. Notice how the meaning indeterminist thesis I have defended thus far is almost identical to this one: Davidson is

¹⁵⁷ As famously argued in Hume (1739). Kripke briefly makes this same comparison in WRPL, pp. 62-64.

¹⁵⁸ This is also stated in Kusch (2006), pp. 83-84.

¹⁵⁹ Sultanescu & Verheggen (2019), p. 24.

simply more comfortable stating that the meaning of words does exist in some sense, and I do not see a reason to reject this proposal. Notice further that there *is* an important difference between the account defended in these pages and Davidson's: as with Lewis and Grice, there seems to be an unnecessary concern for the parties to be *aware* of the exchange they are in. Unless it is to be taken as deeply unconscious or instinctive, I reject (as explained above) the idea that speakers need to know or believe anything specific about their coordination attempts, about the world or about the interlocutor¹⁶⁰, for their attempts to happen, and for them to be successful. On the contrary, the majority of successful communication practices, in actuality, have none of such philosophical concerns laid out before them.

Davidson also helps the cause of meaning indeterminism when he discusses the notion of "idiolects"¹⁶¹. An idiolect is the set of linguistic expressions an individual uses to communicate¹⁶²: what is strikingly important about this is that no-one has the same idiolect as anyone else, or, at least, many people within the same linguistic community have considerably different ones. For instance, I (actually) have the bizarre habit of addressing people and animals with made-up names: my friends and family members have been referred to with countless "nicknames" over the course of my life, names that, importantly, sometimes never expand *outside my individual use*. Yet, it takes very little time for my interlocutors to *understand* what object I refer to by the new name, and to triangulate with me on it.

The important point I want to make here is that KW's claim that "A deviant individual whose responses do not accord with those of the community in enough cases will not be judged, by the community, to be following its rules", can be misleading in this sense: it seems to give justification to a popular reading of the meaning indeterminist position, such that we are right when we use a word, because others would also use it in the same context. But this is false: my "deviant" use of names I use to refer to my friends and family never endangered my role as a member of the linguistic community, even though nobody else was disposed to use those same words. Clearly, what we should take in serious consideration from KW's claim is that we need to accord with what the community would do *in enough cases*: my interlocutors understand what I mean by my made-up nicknames because they are arranged in familiar linguistic contexts: if I say "X went grocery shopping", my interlocutors will easily coordinate with me even if they would never use the word X to refer to the individual who went grocery shopping, but that is because the rest of the sentence is perfectly familiar to them. What is really needed to communicate,

¹⁶⁰ As world and *as* interlocutor. Of course, there are pieces of information *about* the world and the interlocutor that are needed for the communication to work out.

¹⁶¹ Such as in his (1986).

¹⁶² As usual, the best understanding of the notion of "idiolect" is controversial among philosophers. Some go as far as to argue that idiolects don't properly exist. For the purposes of my considerations, assume the one I am presenting is the correct one.

then, is not an adherence to predetermined linguistic expressions¹⁶³, but the ability to coordinate one's attention and someone else's on an object, by whatever means necessary. This, I claim, is equivalent to what Davidson calls triangulation and to what Skyrms, as we have seen, calls a "salient coordination equilibrium". Both notions, in fact, focus on some natural connection we have with our environment and with our peers, due to which we instinctively prefer one object as our center of attention than another when our peers lead us there with their words and signals.

Does the community of speakers have any power over meaning?

Now that I have laid out what I take to be a satisfying answer to KW's question, it is time to come back to his own answer and see where it differs and what I find to be wrong with it. I have claimed, in Section 1, that the first problem with it is the assumption that, due to the argument against the existence of fixed meanings, KW can't but explain his solution in unsystematic terms. My position on this is that one such approach only enhances the shroud of mystery that surrounds much of Wittgenstein's works and that makes it so hard to pin down many philosophers of the past. While this stylistic approach has had its theoretical reasons for some authors (say, Nietzsche), analytic philosophy usually prides itself with being concise and to the point, leaving literary drives to literature. Kripke does seem self-conscious of this choice and argues against the possibility of a clear-cut answer from KW¹⁶⁴, but then again he also claims, as we have seen a few pages ago, that we *can* account for meaning-talk, though without ontologically solid meanings. Though it may be seen as a strong claim, I suspect that all the vagueness in KW's answer is simply the symptom of not having a good answer to his own problem. Let us deal with it.

KW is commonly understood as giving a "social" answer to his own problem, which seems to be the correct reading, as suggested by passages such as: "Someone [is] said to follow a rule, as long as he agrees in his responses with the [...] responses produced by the members of [his] community" and "We say of someone else that he follows a certain rule when his responses agree with our own and deny it when they do not". What this means is, basically, that we understand how to use a word if we do as others -specifically, the linguistic community- would do. Before addressing this idea, we should settle the problem of factualism we introduced in Section 1: is KW offering an account where we are able to determine the truth value of a sentence such as "Joe means *banana* by 'banana'"? I don't believe there is conclusive proof in either direction, so the question might be better answered in these terms: which is the theoretically more constructive way to understand KW's solution?

¹⁶³ If the first understanding of the social claim were true, as I will focus on later, that would imply the impossibility of change in the linguistic habits of a community.

¹⁶⁴ WRPL, p. 69: "Wittgenstein's professed inability to write a work with conventionally organized arguments and conclusions stems at least in part [...] from the nature of his work".

Say the factualist view is. In this case, we need to find a matter of fact that should determine whether a metalinguistic statement is true. In KW's terms, it seems that the only relevant fact could be the adherence to one's linguistic community's usage of the same word. It would be a mistake to think of this adherence in dispositional terms, though: Kripke explicitly rejects this possibility, and for good reason¹⁶⁵. The adherence we are looking for would be better understood in terms of *agreement*: Joe means *Banana* by "banana" iff the way he uses the word is approved sufficiently enough by his linguistic community.

Now, if this is to be taken as a generalized statement, it is, in my view, strictly false. There are at least two good reasons to think that. The first one is to point out that moving the burden of meaning correctness from the individual to a group of individuals, large as it may be, in no way solves the problem KW started from: how is a group more competent about the usage of a word, if we can always, in principle, doubt that the future usage should comply with what we expect from it? In other words, there does not seem to be any important logical difference, when the "paradox" is taken into account, between a single person using a word in such and such way and a group of people doing the same thing. Now, to be fair, this objection lies in the realm of the normative: what KW wants to say seems to be that this is not in principle a good way to determine the meaning of a word, but it is how it happens; after all, he does claim that "The regularity [of linguistic practices] must be taken as a brute fact" (p. 98).

Thus, we must acknowledge that KW understands his solution as giving "*assertability conditions*" rather than correctness conditions, that is, "conditions in which the participants in the relevant practice are disposed to utter it assertively" (Miller 2020, p. 9). As such, it seems more beneficial to look at the answer in descriptive terms: the community of speakers doesn't determine the way we ought to use a word, but it determines the way we actually use it.

It seems to me that the social answer does not work under a descriptive point of view either. What this account fails to explain is, simply, how linguistic drifts from the established use can happen as frequently as they do. I suspect KW's reply to this would sound something like: "Language shifts happen blindly: for some reason that lies beyond the realm of justification, the majority starts to use words in new ways and that changes the social rule of assertability". Again, this does not seem true. Especially in modern times, every change in a language or in the expression of a conceptual framework seems to *demand* a justification. A prominent example of this can be found not too far from this essay itself: in philosophical works, conceptual analysis has at its core the project of understanding where it is *best* to draw the line for some concepts, which are associated

¹⁶⁵ "[A social version of the dispositional theory] would be open to at least some of the same criticisms as the original form".

with words. As Davidson (2001a) underscores: “Simply adding further creatures with identical dispositions cannot turn dispositions into rule-following [...]. What is missing is [...] understanding”.

As brilliantly put by Skyrms & Barrett in their recent (2019), “‘Wolf’ does not get its meaning from all its contexts of use, such as the one that includes the little boy crying wolf¹⁶⁶. This would be true even if crying wolf became more common than in the story. ‘Louis Vuitton’ does not get its meaning from the vendors selling fake plastic handbags, *although they may account for most of the usage*”¹⁶⁷. But then, what is it that gives “Louis Vuitton” its meaning? My take is that, coherently with the story we have been recounting, there is something in the psychology of speakers that makes it more attractive for them to call a real Vuitton bag “Louis Vuitton” rather than a counterfeit one, upon reflection. Skyrms & Barrett seem to agree, and call this a “context of common interest”. I see this as a specific case of what we have introduced as salience in communication.

Basically, when we have to identify the meaning¹⁶⁸ of a word, we shouldn’t look for the way the linguistic community uses it: as we have noticed, that returns us to the dispositional problems the individual, or smaller groups of speakers, faced.

Another sense of “primitivity”: the salience of salient equilibria

We now have every tool at our disposal to understand what should *really* be called “primitive” in the context of communication. As Guardo (2018; 2019) stresses, the essential feature for this practice to work is our “common animal nature”. Since we have many psychological commonalities as members of the same species, we intuit what others want us to direct our attention to (or to triangulate on) when they make a specific sound, not dissimilarly from other animal species: “this seems to be sufficient to conclude that [we] understand each other and, therefore, that [we] manage to communicate” (Guardo 2019, p. 122). The primitive aspect of communication should thus be found in the way we naturally refer to objects in the world; since we are similar under many aspects, we manage to work out most communicative attempts. Notice that this is something Wittgenstein does hint at multiple times: in his *Philosophical Investigations* §140, he argues that what makes it the case that we imagine we have but one choice of application of a word is the fact that only one application comes to mind; in §185, he says that someone who follows a numeric succession in the same way we would do “grasps” the rule because he is naturally led to do so. After all, it seems we always had the answer in front of us, and

¹⁶⁶ Famously, the story of the little boy crying wolf says that a kid that always cries for help when wolves are not really there, will not be saved when the wolves really come. This implies, as a figure of speech, that when someone is “crying wolf”, the wolf is not actually there.

¹⁶⁷ My emphasis.

¹⁶⁸ I take it that, at this point in the essay, I can talk of “meanings” without being accused of falling for the primitivist objection: meaning has turned out to be best understood as a use we find psychologically attractive or appropriate, and that is how I *mean* it, both in metalinguistic and in basic sentences.

that Kripke's interpretation of Wittgenstein's (not too explicit) solution has created more disarray than there would have been without it.

I would like to also point out that our psychological *differences* are as essential as our similarities to understand what happens when we communicate: returning to my "weird naming" example, some people struggle more -or are almost unable- to coordinate with me in such cases, because they are more firmly attached to social conventions or to what we have come to understand as the quasi-realist belief we have about the existence of meanings, for instance. In those cases, an environmental feature in my interlocutors' life experience makes their psychological traits too different from mine in order to successfully use my "creative" naming system to coordinate with them. This can pull me in two different directions depending on what I care about the most: if I enjoy my idiolect more than I care to coordinate with such individuals, I may decide I will not coordinate with them; if I care more about coordinating, I will need to accommodate my idiolect to their expectations. This process happens in all linguistic communities and it is most likely the reason for language shifts¹⁶⁹.

We have first introduced the notion of a "language shift" when addressing the problems KW's solution has to face: the reader will recall that I took the implausibility of one such event in KW's account as a consequence of the supposed need to comply with the other speakers' expectations, as if the correctness of an expression was to be found in what other people *think* the correct use of the expression is. Let us now delve into this idea with more attention to its evolutionary implications. First, we need to distinguish, following Soysal (forth., p.9), a complete resistance to linguistic changes prescribed by external influences and a *partial* one. I take KW's account to support the idea of pure descriptivism, such that the correctness of application of a word is fully determined by what the community thinks it should be. Jackson's (2004) notion of *resilientism* is more helpful to properly describe human communication: *some* aspects of our linguistic uses are simply harder to modify than others. This is perfectly physiological, as implicitly¹⁷⁰ noted in Jackson (1998): there could not be a language shift without some language "stagnation". The basic idea here is that, in order for change to happen, one must start from what already exists. Given this, we could not talk of having a language if, in most cases, our linguistic responses to

¹⁶⁹ A very common example is the progressively popular use of anglophone expressions in non-anglophone communities: such expressions regularly enter the vocabulary of these communities when the majority of the population uses them to coordinate with others, so that even the members of the community who are not particularly disposed or happy to use them are "forced" into the new convention by their need to coordinate with others.

¹⁷⁰ Specifically, pp. 31-33. In these pages Jackson gives a metaphilosophical evaluation of intuition-handling when philosophers address them to make corrections to them. I think the same idea applies to the way we use words, since there is a direct correlation (as usual, in psychological terms) between such words and the concepts in our minds: following his example, once Gettier convinces us that knowledge is not true justified belief, we will not sincerely or correctly use the corresponding words as synonyms.

external impulses were not usually the same over time; apart from being a logically stringent point, this is also evolutionarily justified.

As Skyrms explains¹⁷¹, salient equilibria are usually safe from sudden practical¹⁷² changes (such as a Goodman-like scenario) when we have reached an equilibrium that “is maintained because a *unilateral* deviation makes everyone strictly worse off”. Be reminded that, on paper, equilibria are all equal, as long as no party could have acted more optimally, given how the other parties acted; a salient equilibrium is an equilibrium that is especially attractive for some psychological reason. Clearly, a language shift needs some change in the use of words: but that is sub-optimal in most situations, because, when others are not as prone to that linguistic change, the change does not happen. One simple way for others to accept the linguistic shift is for it to be more optimal than the state-of-the-art linguistic use. Since there is more than one coordination equilibrium, though, it is not clear at first glance what makes one of them be *so* consistently followed by all speakers or, at least, by all *successful* speakers: it would seem that there should be a greater level of disagreement on how to use our words.

Skyrms (2014, pp. 92-95) gives a clear explanation of this phenomenon by looking at a hypothetical scenario, where two signaling system strategies have taken over exactly half of a linguistic community each: we could say that the population’s reason for salience (the reason they chose a specific equilibrium over the others) is perfectly split in two. “Each system has the same average payoff of $\frac{1}{2}$, so the dynamics [...] does not move the population proportions”. We are again in the presence of a logical-psychological dichotomy: both equilibria are logically equivalent, because *if* they are instantiated they give the best possible outcome for all parties to the communication. But, in actual circumstances, it is impossible for this exact division to hold for long: “If [...] there is any noise in the system, the mixed equilibrium will not survive and one or the other of the signaling systems will be selected”. A bit surprisingly, Skyrms adds that “Which [signaling system] will be selected is a matter of chance”. I take this to be an observation in terms of a schematic *simulation* of equilibria: in a natural environment, at least some of the equilibria that are chosen are not merely a matter of chance, because the salience can usually be explained in terms of its influence on the speakers, be it for visual, imaginative, auditory or other aspects of it.

In particular, as Skyrms explains, there is one specific signaling system (the one that is instantiated in actuality) that has “some psychologically compelling quality that attracts the attention of the decision maker”¹⁷³. In the context of a linguistic setting, one could easily produce endless examples of coordination equilibria: for any object x the sender

¹⁷¹ Skyrms (2014), pp. 84-85.

¹⁷² The specification is needed because KW could obviously object that there never was a *change* in the first place and that we have always been following the Goodman-like rule.

¹⁷³ *Ibi.*

wants to refer to, if both the sender and the receiver use the signal y to refer to x , then we have a coordination equilibrium. Interestingly, the variable can *in principle* be replaced by any possible utterance, at least as long as it doesn't clash with similar or identical ones already in use to refer to other objects: at this level of abstraction, words (and other signals) are simply *codes* whose job is to send a message. Thus, in a theoretic schema such as Lewis's, any word would be on the same level as any other as a candidate for the coordination equilibrium to happen; on the other hand, in the actual conditions where the coordination equilibrium is being sought by the speakers in a natural environment, some words will be chosen more likely than others, for reasons that fall outside the realm of semantics: semantically speaking, then, these preferences can genuinely be said to be primitive.

These psychologically compelling qualities are not necessarily justified in a strong sense: they may be subjectively preferred by the leader of the bunch, so that everyone else is compelled to follow their example for "political" reasons; they might sound better as a metaphorical beacon for the referent (what is usually called an "onomatopoeic" name); it might just be, as for the majority of the words we use, what the society we live in takes as the "correct" name of an object¹⁷⁴, and thus it is introduced to us as such as children. What really matters is that, *whatever the reason*, there is always one of the infinitely many¹⁷⁵ viable options that "stands out" more than any other in our minds: since we are similar, fortunately we usually find ourselves agreeing on which one it is. We have called this coordination equilibrium we naturally drift towards, "salient equilibrium", after Lewis and Skyrms: this is the key to understanding why both linguistic stasis and linguistic shifts can happen. Linguistic stasis happens when no symbol which is being used by deviant speakers has more psychological strength than the symbol already in use for most of the community's communicative interactions: in such scenarios, the deviant speakers simply stop using the deviant symbol because it impedes their coordination attempts with others. Linguistic shifts happen when one of the symbols used by a deviant speaker gains more psychological strength than the symbol already in use, thus "dethroning" it.

Now, it seems to me that this is where we can shut off Hattiangadi's worry: the existence of salient equilibria makes it the case that, although *in theory* one could use or understand words differently than most speakers do, *in reality* this cannot really happen, as long as the current salient equilibrium is not replaced by a new one¹⁷⁶, due to our common biological nature. Now, what we took to be the really worrisome aspect of Hattiangadi's remarks was that meaning indeterminism can't seem to account for the *preciseness* we expect from

¹⁷⁴ Of course, this social preference is itself caused by other salient reasons and is to be taken as primitive *only from the perspective of the new member of the semantic community*.

¹⁷⁵ Since, in principle, any referent could be attributed any word/symbol with infinitely many letters/elements.

¹⁷⁶ Where, of course, the new salient equilibrium is still to be justified in psychological terms: it is simply the word/symbol the *actual* creatures that communicate find most appealing for the job, for some reason.

the employment of our words. Granted that there is always some degree of deviation, well captured by the notion of idiolects, for the most part, for the purposes of communication, such deviations don't compromise the understanding required of the receiver by the sender in the vast majority of cases. Let us see this idea with an image:

Assume there are 100 objects in the world, such that some of them are cats and some are not. John uses the expression "cat" expecting to direct his interlocutor's attention to his cat. This interaction is most likely to succeed, because, out of every object in the world, John and his interlocutor psychologically assign *Cat* or *non-Cat* in the same way to 99 out of 100 of them; that is to say, their idiolects are extremely similar. The one object John would *be moved towards* when hearing the word "cat" and his interlocutor would not (or vice versa) is the deviation from their common understanding of the word "cat", that is, what makes John's idiolect his own idiolect and not his interlocutor's (and vice versa), which disrupts their chances of triangulation should they find themselves in the situation. But, importantly, John *can* convince his interlocutor (or vice versa) that the one object they disagree on ought to be called in the way John would call it: not because of some rule *Cat* somewhere, that is supposed to pre-determine the correctness of any possible use of the word "attached" to it, but because their environment, their history with the usage of the word "cat" and of other words, and the results of conceptual engineering¹⁷⁷, can psychologically convince one of the speakers to behave as the other does. A very simple, but common and objectively useful, example would be for John to point out that *the community of speakers* triangulates on that object with that name: John's interlocutor would most likely be moved, then, to accept such usage of the word, because otherwise he would find himself in the risky position of being the deviant speaker, unable to coordinate not only with John, but with *anyone*.

My last example was, as the reader may have noticed, intended as a final rebuttal to KW's own defense of meaning indeterminism: the social answer turns out to be a local case of the general idea of salient equilibria in communication and, in particular, in the way we use our words, in language.

Final words

Hopefully, the account of communication I presented in this essay has given reason to be less distrustful of the meaning indeterminist proposal. The recent literature on Kripkenstein's problem has, in fact, shed light on the fact that effective communication is possible without fixed meanings, and the recent developments in evolutionary studies have shown that the animal kingdom shares countless examples of this with humankind. These contributions turned the plausibility of meaning determinism on its head: to

¹⁷⁷ "Conceptual engineering" is a relatively new notion in metaphilosophy, which suggests that philosophical analysis should look at how a concept *ought* to be used instead of simply describing the common-sense use we make of it.

imagine that one specific animal species, us humans, should have the peculiar power to identify rules fixed in stone somewhere when we use our words is, at the very least, more in need of further explanations than the opposite stance is.

Furthermore, the main goal of this work was to propose an outlook that doesn't outright reject the meaning determinist solutions, but that contextualizes them in a world that finds no *metaphysical* need for semantic meanings. If anyone needs them, it is us: we, as communicators, prefer to attach our words to rules that we fictionally locate somewhere out there, in the same way that we may be said to apply a fictional filter to many other aspects of our lives, such as death or coincidences. Just like those other fictional aspects of our lives, there is no need for us to be *conscious* of the fictionalization and there is no reason to consider such fictionalizations negatively *per se*. At least in the context of communication, in fact, the fictional existence of meanings (in the "strong" sense) seems very useful for coordinating better and faster: thus, there is no need to break the spell, so to speak.

That said, as philosophers we usually want to see the truth as well: at the very least, in order to be able to determine whether the fictional objects we use in our meaning-talk *are* useful, or if they may need changing. After an assessment of the contemporary positions in the debate, it seems realistic to accept meaning indeterminism as the correct position both in metasemantics and in the philosophy of mind, especially since we have shown that it puts the possibility of successful communication in no danger whatsoever.

If primitivism teaches us that we (ought to) make-believe that our words have a set-in-stone meaning, dispositionalism teaches us the opposite: the raw truth of communication is that we are just animals, though very psychologically complex ones, and as such we use sounds and other stimuli to behave and make others behave in ways we desire. Thus, the account I presented here should sound palatable to any meaning determinist theorist who can accept the removal of meanings from our ontology given a consistent framework for communication to work, and who is comfortable with giving new life to pre-existing theories within that same framework, instead of just leaving them to the side.

Though consistent, this account is, obviously, not complete. If my thesis is correct, the next steps in the right direction should be, in my view, concerned with integrating all other plausible meaning determinist answers that have been developed through the years inside the naturalistic framework, by finding what (if anything) they tell us about our psychological lives in communication, similarly to what I tried to do with dispositionalism and primitivism. As I briefly hinted at in these pages, I am confident this project can be

fulfilled in regard to Reference Magnetism and the Simplicity answer¹⁷⁸ (which is, in fact, related to Magnetism), but it is likely that this is still not all there is to it.

¹⁷⁸ In particular, I think that the key to integrating these further answers should be, once again, found in the *naturalness* of some word usages rather than others. What is *wrong* with these answers is that such naturalness is not due to the existence of corresponding universals (as Lewis 1983 thinks) or to the higher simplicity of one rule over another (there seems to be no way to assess objectively whether addition or quaddition is “simpler”), but to the salient equilibria that emerge from our environment.

Bibliography

Azzouni, Jody, *Semantic Perception: How the Illusion of a Common Language Arises and Persists*, Oxford University Press USA, 2013.

Båve, Arvid, 'Semantic dispositionalism without exceptions', in *Philosophical Studies*, Vol. 177, n° 6, 2020, pp. 1751-1771.

Bennett, Karen, 'By Our Bootstraps', in *Philosophical Perspectives*, Vol. 25, n° 1, 2011, pp. 27-41.

Blackburn, Simon, *Spreading the Word*, Clarendon Press, 1984a.

Blackburn, Simon, 'The Individual Strikes Back', in *Synthese*, Vol. 58, n° 3, 1984b, pp. 281-302.

Boghossian, Paul, 'The Rule-Following Considerations', in *Mind*, Vol. 98, n° 392, 1989, pp. 507-549.

Bourget, David & Chalmers, David John, 'What do philosophers believe?', in *Philosophical Studies*, Vol. 170, n° 3, 2014, pp. 465-500.

Chalmers, David John, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, 1996.

Crane, Tim, *Elements of Mind. An Introduction to the Philosophy of Mind*, Oxford University Press, 2001.

Davidson, Donald, 'First Person Authority', in *Dialectica*, Vol. 38, n° 2/3, 1984, pp. 101-111.

Davidson, Donald, 'A Nice Derangement of Epitaphs', in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Ernest Lepore ed., Blackwell, 1986, pp. 433-446.

Davidson, Donald, 'The Second Person', in *Midwest Studies in Philosophy*, Vol. 17, 1992, pp. 255-267.

Davidson, Donald, 'Indeterminism and Antirealism' in *Realism/Antirealism and Epistemology*, C. B. Kulp ed., Rowman & Littlefield, 1997, pp. 109-122.

Davidson, Donald, 'Externalisms', in *Interpreting Davidson*, Petr Kotatko, Peter Pagin & Gabriel Segal eds., Stanford University Center, 2001a, pp. 1-16.

Davidson, Donald, *Subjective, Intersubjective, Objective*, Clarendon Press, 2001b.

Dennett, Daniel, 'The Unimagined Preposterousness of Zombies', in *Journal of Consciousness Studies*, Vol. 2, 1995, pp. 322-326.

Di Martino, Carmine, *Viventi umani e non umani*, Edizioni Libreria Cortina Milano, 2017.

- Douglas, Samuel Paul, 'The *Qua*-Problem and Meaning Scepticism', in *Linguistic and Philosophical Investigations*, Vol. 17, 2018, pp. 71-78.
- Fine, Kit, 'Guide to Ground', in *Metaphysical Grounding*, Fabrice Correia & Benjamin Schnieder eds., Cambridge University Press, 2012, pp. 37-80.
- Fodor, Jerry, *A Theory of Content and Other Essays*, MIT Press, 1990.
- Ginsborg, Hannah, 'Primitive Normativity and Skepticism about Rules', in *Journal of Philosophy*, Vol. 108, n° 5, 2011a, pp. 227-254.
- Ginsborg, Hannah, 'Inside and Outside Language: Stroud's Nonreductionism about Meaning', in *The Possibility of Philosophical Understanding: Reflections on the Thought of Barry Stroud*, Jason Bridges, Niko Kolodny & Wai-Hung Wong eds., Oxford University Press, 2011b, pp. 147-181.
- Godfrey-Smith, Paul, 'Review of Brian Skyrms' *Signals*', in *Mind*, Vol. 120, 2012, pp. 1288-1297.
- Greenberg, Alexander, 'Should I believe all the truths?', in *Synthese*, Vol. 197, no. 8, 2020, pp. 3279-3303.
- Grice, Paul, 'Meaning', 1957, in his *Studies in the Way of Words*, Harvard University Press, 1991.
- Guardo, Andrea, 'Kripke's Account of the Rule-Following Considerations', in *European Journal of Philosophy*, Vol. 20, n° 3, 2012, pp. 366-388.
- Guardo, Andrea, *Seguire una regola*, Mimesis Edizioni, 2018.
- Guardo, Andrea, 'Yet Another Skeptical Solution', in *Philosophia*, Vol. 47, n° 1, 2019, pp. 117-129.
- Guardo, Andrea, 'Two Epistemological Arguments against Two Semantic Dispositionalisms', in *Journal for the Philosophy of Language, Mind and the Arts*, Vol. 1, n° 1, 2020a, pp. 1-13.
- Guardo, Andrea, 'Meaning relativism and subjective idealism', in *Synthese*, Vol. 197, n° 9, 2020b, pp. 4047-4064.
- Guardo, Andrea, 'Yet Another Victim of Kripkenstein's Monster: Dispositions, Meaning and Privilege', forth.
- Hattiangadi, Anandi, *Oughts and Thoughts: Rule-Following and the Normativity of Content*, Oxford University Press, 2007.
- Hattiangadi, Anandi, *The Return of Kripkenstein*, manuscript, 2019.

- Hattiangadi, Anandi, 'No, one should not believe all truths', in *Inquiry: An Interdisciplinary Journal of Philosophy*, Vol. 62, n° 9-10, 2019, pp. 1091-1103.
- Heil, John & Martin, C.B., 'Rules and Powers', in *Philosophical Perspectives*, Vol. 12, n° 1, 1998, pp. 283-312.
- Hume, David, *Trattato sulla natura umana (A Treatise of Human Nature)*, Laterza, 1982 [1739].
- Jackson, Frank, *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford University Press, 1998.
- Jackson, Frank, 'Why we need a-intentions', in *Philosophical Studies*, Vol. 118, 2004, pp. 257-277.
- Jacob, Pierre, 'Intentionality', Stanford Encyclopedia of Philosophy, 2019 (<https://plato.stanford.edu/entries/intentionality/>).
- Kearns, Stephen & Magidor, Ofra, 'Semantic Sovereignty', in *Philosophy and Phenomenological Research*, Vol. 85, n° 2, 2012, pp. 322-350.
- Khani, Ali Hossein, 'Review of Donald Davidson's *Triangulation Argument: A Philosophical Inquiry*' in *International Journal of Philosophical Studies*, Vol. 26, 2017, pp. 113-117.
- Khani, Ali Hossein, 'Kripke's Wittgenstein's Sceptical Paradox: A Trilemma for Davidson', in *International Journal for the Study of Skepticism*, Vol. 9, n° 1, 2019, pp. 21-37.
- Kirkham, Alexander, Breeze, Julian & Marí-Beffaa, Paloma, 'The impact of verbal instruction on goal-directed behaviour' in *Acta Psychologica*, Vol. 139, n° 1, 2012, pp. 212-219.
- Kripke, Saul A., *Naming and Necessity*, Harvard University Press, 1980.
- Kripke, Saul A., *Wittgenstein on Rules and Private Language*, Blackwell Publishing, 1982.
- Kusch, Martin, *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein*, McGill-Queen's University Press, 2006.
- Lewis, David, *Convention*, Harvard University Press, 1969.
- Lewis, David, 'New Work for a Theory of Universals', in *Australasian Journal of Philosophy*, Vol. 61, n° 4, 1983, pp. 343-377.
- Lewis, David, *On the Plurality of Worlds*, Wiley-Blackwell, 1986.
- Mackie, John Leslie, 'Self-Refutation – A Formal Analysis', in *Philosophical Quarterly*, Vol. 14, n° 56, 1964, pp. 193-203.
- Malpas, Jeff, 'Donald Davidson', Stanford Encyclopedia of Philosophy, 2019 (<https://plato.stanford.edu/entries/davidson/>).

- McDowell, John, *Mind, Value & Reality*, Harvard University Press, 1998.
- Miller, Alexander, 'Rule-following, Meaning, and Primitive Normativity', in *Mind*, Vol. 128, n° 511, 2019, pp. 735-760.
- Miller, Alexander, 'What is the Sceptical Solution?', in *Journal for the History of Analytical Philosophy*, Vol. 8, n° 2, 2020, pp. 1-22.
- Ramirez Calle, Olga, 'Meaning under the Threat of Paradox on Two Fronts', in *Analiza i Egzystencja*, Vol. 50, n° 1, 2020, pp. 5-17.
- Sandler, Sergei, 'Fictive interaction and the nature of linguistic meaning', forthcoming in *The conversation frame: Forms and functions of fictive interaction*, Esther Pascual & Sergei Sandler eds., John Benjamins Publishing.
- Schaffer, Jonathan, 'On What Grounds What', in *Metametaphysics: New Essays on the Foundations of Ontology*, David Manley, David J. Chalmers & Ryan Wasserman eds., Oxford University Press, 2009, pp. 347-383.
- Searle, John, *Seeing Things as They Are: A Theory of Perception*, Oxford University Press, 2015.
- Shafiei, Mohammad, *Meaning and Intentionality. A dialogical approach.*, London: College Publications, 2018.
- Shroeder, Tim, Stanford Encyclopedia of Philosophy, 2015 (<https://plato.stanford.edu/entries/desire/>).
- Sillari, Giacomo, 'Rule-Following as Coordination: A Game-Theoretic Approach', in *Synthese*, Vol. 190, n° 5, 2013, pp. 871-890.
- Skyrms, Brian, *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press, 2004.
- Skyrms, Brian, *Signals: Evolution learning and information*, Oxford University Press, 2010.
- Skyrms, Brian, *Evolution of the Social Contract*, Cambridge University Press, 2014 (Second Edition) [1996].
- Skyrms, Brian & Barrett, Jeffrey, 'Propositional content in signals', in *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 74, 2019, pp. 34-39.
- Sládeček, Michal, 'What does the 'bedrock' of rules consist of?: McDowell on Kripke's and Wright's interpretation of Wittgenstein' in *Theoria*, Vol. 55, n° 4, 2012, pp. 5-20.
- Soames, Scott, 'Skepticism about Meaning: Indeterminacy, Normativity and the Rule-Following Paradox', in *Canadian Journal of Philosophy*, vol. 27, n° 1, 1998, pp. 211-249.

- Soysal, Zeynep, 'From metasemantics to analyticity', in *Philosophy and Phenomenological Research*, forth.
- Stoljar, Daniel & Damnjanovic, Nic, 'The Deflationary Theory of Truth', Stanford Encyclopedia of Philosophy, 2010 (<https://plato.stanford.edu/entries/truth-deflationary/>).
- Sultanescu, Olivia & Verheggen, Claudine, 'Davidson's Answer to Kripke's Sceptic', in *Journal for the History of Analytical Philosophy*, Vol. 7, n° 2, 2019, pp. 7-28.
- Tomasello, Michael, *A Natural History of Human Thinking*, Harvard University Press, 2014.
- Verheggen, Claudine, 'Donald Davidson: Looking Back, Looking Forward', in *Journal for the History of Analytical Philosophy*, Vol. 7, n° 2, 2019.
- Walton, Kendall, *Mimesi come far finta (Mimesis as Make-Believe)*, Mimesis Edizioni, 2011 [1990].
- Warren, Jared, 'Killing Kripkenstein's Monster', in *Noûs*, Vol. 54, n° 2, 2020, pp. 257-289.
- Watzlawick, Paul, 'Some Tentative Axioms of Communication' in *Pragmatics of Human Communication*, Donald deAvila Jackson, Janet Beavin Bavelas & Paul Watzlawick eds., 1967.
- Wheeler, Brandon, 'Monkeys crying wolf? Tufted capuchin monkeys use anti-predator calls to usurp resources from conspecifics', in *Proceedings of the Royal Society of London B: Biological Sciences*, Vol. 276, n° 1669, 2009, pp. 3013-3018.
- Wittgenstein, Ludwig, *Ricerche filosofiche (Philosophical Investigations)*, Einaudi, 2014 [1953].
- Wittgenstein, Ludwig, *Remarks on the Foundations of Mathematics*, MIT Press, 1983 [1956].
- Wittgenstein, Ludwig, *Della certezza (On Certainty)*, Einaudi, 1999 [1969].
- Zalabardo, José, 'Kripke's Normativity Argument', in *Canadian Journal of Philosophy*, Vol. 27, n° 4, 1997, pp. 467-488.