The Bivalent Trap
Vagueness, Theories of Meaning, and Identity

MSc Thesis *(Afstudeerscriptie)*

written by

José Pedro Correia
(born February 4th, 1984 in Braga, Portugal)

under the supervision of **Dr. Michael Franke** and **Dr. Ing. Robert van Rooij**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
August 28, 2013	Dr. Maria Aloni
	Dr. Raquel Fernandez Rovira
	Dr. Michael Franke
	Dr. Ing. Robert van Rooij

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

The more narrowly we examine actual language, the greater becomes the conflict between it and our requirement. (For the crystalline purity of logic was, of course, not something I had discovered*: it was a requirement.) The conflict becomes intolerable; the requirement is now in danger of becoming vacuous.—We have got on to slippery ice where there is no friction, and so, in a certain sense, the conditions are ideal; but also, just because of that, we are unable to walk. We want to walk: so we need* friction. *Back to the rough ground!*

– Ludwig Wittgenstein

Abstract

In this thesis we explore the implications of vagueness for theories of meaning. We argue that vagueness in natural language presents a serious challenge to the traditional conception of meaning as a relation of determinate correspondence between signs and entities. The way forward, we defend, is to view meaning as an emergent property of a complex system. In search of approaches that allow for a formal grip on meaning within this perspective, we turn to the framework of signaling games. We demonstrate the advantages of the change in paradigm by providing examples of how one can model vagueness and the sorites paradox in a manner that allows us to study the phenomena transparently, rather than obscuring them or explaining them away. Finally, we illustrate the generality of the approach by applying it to a related philosophical problem: identity and change.

Contents

Acknowledgments

I would like to thank Michael Franke and Robert van Rooij for agreeing to supervise my thesis, the time they put in for meetings and reviewing my work, and all the useful comments, feedback, and discussions throughout this past semester. I would also like to thank Martin Stokhof for a very interesting conversation on vagueness, correspondence, and Wittgenstein. His comments helped shape the argumentation in chapter 4, although he might not necessarily endorse my own opinions and conclusions. Finally, I would like to thank Elliott Wagner for reviewing a first draft of my thesis and for very helpful comments.

Certain people have been instrumental in shaping the academic process that led to this thesis. Within the MSc in Logic, I would like to thank Elsbeth Brouwer for the two courses in philosophy of language I took with her, the first of which was responsible for sparking my interest in the topic and leading me to pursue that research direction, the second was where some ideas in this thesis started taking shape. I am also grateful to Michael Franke and Elliott Wagner for organizing a very inspiring course on language and game theory, which was paramount in laying the ground for the work in this thesis. Thank you also to my fellow students for interesting discussions and collaborations for different courses, and for making the whole experience also a socially gratifying one! Finally, I would like to thank José Nuno Oliveira and Joost Visser, not only for the recommendation letters that enabled my application to the MSc in Logic, but also for everything they taught me.

On a personal note, I would like to thank all my friends for helping me keeping my sanity during stressful times by joining me for drinks, pool, concerts, poker games, football, or anything else to keep my mind away from philosophy for a while. I am also very grateful to Tamara for all of that, and for her love, patience and support! Finally, I want to thank my parents for all of their support throughout my life. Mãe, pai, obrigado por todo o apoio que sempre me deram, por sempre acreditarem em mim e me incentivarem a seguir o meu caminho inclusivé na Holanda, apesar da saudade e da distância.

Chapter 1

Introduction

The amount of confidence we deposit in our natural language is quite remarkable. We use it to get through the day, whether it is to buy a cup of coffee, to ask someone for information on the street, or to arrange a meeting with some friends. We use it at work, to communicate with colleagues, exchange ideas, or just talk about our weekend. We use it to make important decisions, be it in a meeting to decide layoffs in a company, in government to plan economic policies that might affect a great deal of the population, or in a courtroom to decide if someone committed a crime and must be punished for it. Not only that, but we use it constantly in our scientific endeavors to gain knowledge about the world. We use it to write articles and books, to discuss results in conferences and panels of experts, to devise scientific standards, and to teach others what we know. Even in such an abstract area as mathematics, we rely on natural language to design formal systems, share our intuitions about them and their applications, and convince others of the coherence of theorems and the validity of their proofs.

There is at least one good reason why we rely on natural language to do all these things and more: it works. Although not perfect at it, natural language is quite successful in enabling us to communicate and coordinate. And yet, as soon as we start putting it under the microscope and challenging our pre-theoretical understanding of its workings, we easily stumble upon resilient problems that seem to question its very coherence. The topic of this thesis is one of such apparent problems: vagueness.

It feels uncontroversial to say that there are tall people and those who are not tall. But at which specific height does one go from one category to the other? There are certain piles of sand that are uncontroversially heaps, and those that are not. But what exact number of grains of sand draws the border? We say that these words are *vague* in that they do not seem to have precise definitions that allow us to fully determine all the cases in which they clearly apply or not. This, *prima facie*, seems to imply that we should not claim that either someone is tall or not tall, either something is a heap or not a heap. Our vague words do not seem to warrant us thinking about them in a bivalent manner. However, we typically use them as if they do, even in our scientific endeavors:

> Bivalence is a basic trait of our classical theories of nature. It has us positing a true-false dichotomy across all the statements that we can express in our theoretical vocabulary, irrespective of our knowing how to decide them. [Quine, 1981, p. 94]

The ubiquity of bivalence in our ways of thinking and speaking is even more baffling when we realize, on the other hand, how pervasive vagueness is in the very concepts and words that seem to ground our thought and speech. The two appear to be in sharp conflict with one another, and yet naturally coexist.

The problem of vagueness has led many to dwell on considerations about the nature of reality, knowledge, and language. In chapter 2 we try to characterize the problem in some more detail. We also explore arguments for and against the pervasiveness and importance of the phenomenon. That will lead us to discuss some ontological interpretations, and finally to unearth a shared assumption of these: a picture of natural language where meaning is seen as a form of correspondence between words and entities. This assumption, we believe, is crucial to understanding the problem posed by vagueness. Based on that, we focus, in chapter 3, on the importance of this notion of meaning as correspondence for our philosophical and more formal theories of meaning. We explore how vagueness can be seen to challenge that notion, and question the plausibility of correspondence-based approaches that try to explicitly address the problem of vagueness. We argue that, in the face of vagueness, a new way of conceptualizing meaning is needed to better reflects our intuitions. If we aim to properly understand natural language, we have to acknowledge vagueness as an important characteristic, instead of trying to avoid it, correct for it, or explain it away.

In search for an alternative, in chapter 4 we turn to the idea of conceptualizing meaning in terms of emergence. We first briefly discuss the later philosophy of Ludwig Wittgenstein and, more concretely, what does it mean to conceptualize meaning as use. We argue that the main tenets of this idea can be found in a formal theory of meaning, namely signaling games, and that it amounts to seeing meaning as an emergent property of a complex system that is irreducible to mere correspondence. In chapter 5 we give an example of how vagueness can be accounted for in signaling games. We do this by showing how a vague language can emerge out of some basic low-level constraints, and how such a language can be largely beneficial, despite its vagueness. In line with our promise of taking vagueness seriously, in chapter 6 we show how and under which conditions such a language can be vulnerable to the sorites paradox.

In order to show the benefits of the change of paradigm, in chapter 7 we try to tackle a seemingly different but very related problem, namely that of identity and change. We try to show how a formalization of everyday identity in the same lines as the proposed approach for modeling vagueness can help us explain, not only how we can have a tolerant notion of identity that can be vulnerable to paradoxes similar to the sorites, but also how it can in fact be robust to very similar but naturally occurring situations, such as continuous change.

Chapter 2

Vagueness

Vagueness is a property of words in natural language that is typically characterized in terms of three phenomena:

1. vague words have *borderline cases*, *i.e.* situations where the word seems to neither positively, nor negatively apply;

2. there are *no clear boundaries* discriminating between the cases where the word seems to positively apply, those where it seems to negatively apply, and the borderline cases;

3. vague words are vulnerable to the *sorites paradox*.

We will here focus mostly on the characterization of vagueness in terms of the sorites paradox. The reason is that the paradox serves mainly to illustrate the lack of clear boundaries in the application of a word, which in turn is responsible for the uncertainty in borderline cases.

The original formulation of the sorites paradox is attributed to Eubulides, who lived in Ancient Greece. Despite its antiquity, the paradox continues to challenge philosophers to this day. It goes as follows. Consider a heap of sand in front of you. If you remove one grain of sand from it, the result is still a heap, albeit one grain of sand smaller. Given that what you have in front of you now is still a heap, you can repeat the process, remove another grain of sand, and you will still have a heap in front of you. The paradox arises since, if you hold to the validity of the premises, the process can be repeated until only one grain of sand is in front of you and you would thus be driven, by your reasoning, to still call this one grain of sand a heap. In fact, since that is a heap, you can as well remove one grain of sand from it and you would still have a heap, according to the premises, even though you are left with nothing. Seemingly, the alternative would be to say that at some point in this process the removal of one grain of sand turns a heap into a non-heap, which is quite counter-intuitive. The argument can also be devised in the negative form: one grain of sand is not a heap; given something that is not a heap, stacking one grain of sand on top of it does not make it a heap; therefore, no amount of grains of sand stacked on top of each other can ever be a heap.

In both forms, the paradox seems to expose a clash between our common sense intuitions that (1) a large enough number of grains of sand stacked on top of each other could certainly be a heap, (2) one grain of sand is certainly not a heap, and (3) the adding or removal of one grain of sand certainly does not make a difference towards whether something is a heap or not. The sorites argument simply brings to light that, taken all together with standard inductive reasoning, these three intuitions are logically inconsistent. But how serious is the problem?

2.1 The pervasiveness of vagueness

Most people, if not everyone, would agree that vagueness exists in language. When we speak about vagueness in language, we typically speak about vagueness in the *meaning* of singular terms and expressions. ‘Nice’ is a paradigmatic example of a vague term. Other terms like ‘tall’, ‘bald’, or ‘red’ might require convincing through a sorites argument, but their vagueness is undeniable. How widespread is the phenomenon? Bertrand Russell [1923] famously argued that all natural language is essentially vague:

> The fact is that all words are attributable without a doubt over a certain area, but become questionable within a penumbra, outside which they are again certainly not attributable. Someone might seek to obtain precision in the use of words by saying that no word is to be applied in the penumbra, but unfortunately the penumbra itself is not accurately definable, and all the vaguenesses which apply to the primary use of words apply also when we try to fix a limit to their indubitable applicability. [1923, pp. 63–64 (reprint)]

This claim has been restated more recently by Michael Dummet [1975] and other authors, but it is sometimes disputed by others.

Christopher Kennedy [2007] argues that absolute gradable adjectives, like for example ‘wet’ and ‘pure’, are not vague: a table is wet if there is some amount of water on it; a piece of gold is pure if it contains no amount of other materials. We believe this claim does not adequately characterize our everyday use of these terms. We might not necessarily call a table with one drop of water on it ‘wet’ and if a person claims they have a pure gold coin, we might not necessarily interpret that as meaning the coin consists exclusively of gold molecules. Even if we ignore these examples, which Kennedy characterizes as imprecise use of non-vague terms, we need to further consider our perceptual limitations. One cannot visually distinguish a table with one water molecule on it from one without, thus it would be absurd to say that a claim that ‘the table is not wet’ is correct in the latter case but incorrect in the former[1]. These considerations need to be taken into account if our theory of vagueness is to be plausible.

Was Russell thus correct in his assertion that all natural language is vague? Just the fact that we can formulate and debate this question seems to imply that we are able to conceptualize a precise language and thus seems to leave open the possibility of the existence of non-vague terms to support this intuition. The closest thing to this ideal is perhaps to be found in formal systems of logic. Russell, however, believed that even logic is tainted by vagueness by ultimately being grounded in words and concepts from natural language:

> Since all non-logical words have this kind of vagueness, it follows that the conceptions of truth and falsehood, as applied to propositions composed of or containing non-logical words, are themselves more or less vague. Since propositions containing non-logical words are the substructure on which logical propositions are built, it follows that logical propositions also, so far as we can know them, become vague through the vagueness of “truth” and “falsehood”. We can see an ideal of precision, to which we can approximate indefinitely; but we cannot attain this ideal. [1905, p. 65 (reprint)]

It seems from this argument, however, that a formal system of logic (or other mathematical language) could avoid vagueness by being kept absolutely abstract. In order to properly explore this problem, we would need to delve into the philosophy of mathematics, which would substantially sidetrack us from our main topic. Therefore, we will leave room for the possibility of precision, but we will henceforth assume that *the great majority of singular terms and expressions in natural language is vague*. In fact, the phenomenon is so pervasive that it is intimately related to another philosophical problem: identity and change.

2.2 Identity and change

The most paradigmatic formulation of the problem of identity and change is the Ship of Theseus paradox. Its origins go back to the works of Plutarch:

[1] See also another formulation of this point by Rohit Parikh [1994, pp. 522–523].

> The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their pace, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same.[2]

Though the formulation here is less explicit than our formulation of the sorites paradox in the beginning of this chapter, the parallels between the two should be clear: we start with an entity to which a term clearly applies; we modify it step by step in a way that no particular step seems to justify ceasing to apply the term; via this process we arrive at a situation where the entity as changed so much that we would not originally be willing to apply the term, but are nevertheless inclined to, given the gradual nature of the change.

The main difference between the two paradoxes is that, in the Ship of Theseus paradox, both the entity we start with and the entity at the end of the process are still of the same kind, thus focusing the question simply on whether it is *the same* entity or not. However, the same principles are involved in both, namely determinate reference and tolerance under change. The sorites paradox is in fact more general since a question of identity runs parallel to the question of applicability of a term: not only we can ask ourselves when does our heap stop being a heap, but also when does it stop being *the same heap*. For this reason, we will mostly focus on vagueness and the sorites paradox, but in chapter 7 we will come back to the Ship of Theseus and the consequences of our findings on vagueness for the problem of identity and change.

Another important difference between the two paradoxes is that the second is typically considered to expose an ontological problem, whereas the first is usually considered a linguistic problem. However, there is a significant number of authors that interpret the sorites paradox as bearing on reality, rather than merely language.

2.3 Ontological interpretations of vagueness

Peter Unger has an interesting take on the sorites paradox. He claims there is nothing paradoxical about it, in a philosophically important sense. Rather, the sorites argument is simply a sound logical proof that such a thing as a 'heap' does not exist:

> In its original form, the sorites argument appears to have concerned how many items, say beans, or grains of sand, or even some of each, will be sufficient to constitute a heap. None or one is insufficient. But, if there isn't any heap before us adding a single grain or bean, it seems, will not produce a heap, Hence, even with a million beans quite nicely arranged, there will be no heap of them. By generalization, this is a compelling argument that there are no heaps, and that our concept of a heap is relevantly incoherent. It is, we might say, a direct argument for this idea and, I believe, it is a sound one. Conversely, we may begin by supposing that there are heaps, and that a million beans typically arranged gives us an instance of that concept. But, then, removing a single peripheral bean gently from such a typical heap, it seems, will not leave us with no heap before us. Hence, we must conclude that even when we have but one bean left, or none at all, we still have a heap of beans. But this is absurd. Hence, we have reduced the original supposition of existence to an absurdity, and we may generalize accordingly. This, we may say, is an indirect argument that there are no heaps, and that our concept of them is not a coherent one. It is also, I believe, an adequate argument. Now, Eubulides' seminal contribution has long labored under the misnomer of 'the sorites paradox'. But, in any philosophically important sense, there is no paradox here. Rather, we are given two demonstrations of the non-existence of

[2] Plutarch, *Life of Theseus*, verse XXIII. Obtained from http://classics.mit.edu/Plutarch/theseus.html (accessed 15/07/2013).

> heaps, while no important logical problems come from accepting the conclusion. [Unger, 1979, p. 118]

The overarching ontological position, also defended by other authors such as Samuel Wheeler [1975; 1979] and Terence Horgan [1994; 2009] and sometimes known as *mereological nihilism*, defends the idea that the only things that properly exist are necessarily indivisible. The rest, everything that is composed of parts, merely corresponds to different arrangements (spatial, temporal, topological) of the basic constituent(s), without any special ontological status. From this point of view, there is no actual vagueness in the world; vagueness is the result of a mismatch between our presuppositions of ordinary objects with determinate boundaries and how the world actually is.

Another approach that also denies vagueness in the world is the *epistemic view*. Some authors, such as James Cargile [1969], Roy Sorensen [1988], and Timothy Williamson [1992; 1994], defend that ordinary objects do exist and there is a fact of the matter where their boundaries lie. Going back to our example of the heap, this means that there is an objective minimum number of grains of sand that constitute a heap and a pile of sand with one grain of sand below that number is, as a matter of fact, not a heap. Vagueness is the result of our epistemic limitations: we don't know precisely where this boundary lies, and we may not be able to come to know it. The approach confines vagueness to the epistemic realm, but holds to a precise ontology mainly in order to be conservative regarding our current theories of truth and meaning:

> The most obvious argument for the epistemic view of vagueness has so far not been mentioned. The epistemic view involves no revision of classical logic and semantics; its rivals do involve such revisions. Classical logics and semantics are vastly superior to the alternatives in simplicity, power, past success, and integration with theories in other domains. [Williamson, 1992, p. 279 (reprint)]

The epistemic view is therefore not *per se* an ontological interpretation of vagueness, but it does make explicit ontological claims.

Finally, another approach is to consider that vagueness exists in the world, *i.e.* there is actual *ontic vagueness*. Authors such as Peter van Inwagen [1988], Michael Tye [1990], and Eddy Zemach [1991] give arguments for the existence of vague objects. These are entities in the world which have indeterminate spatio-temporal boundaries. A typical illustrative example is a cloud, where the concentration of water molecules drops off from the center in a continuous manner, making it effectively impossible to determine where the cloud ends and the clear sky begins. Vagueness in language is then merely a reflection of vagueness in the world. The idea of vague objects has seen a considerable amount of attack, which has lead some authors to defend different approaches to ontic vagueness. Terence Parsons and Peter Woodruff [1995], for example, explore the idea of vagueness not in objects, but in states of affairs, *i.e.* in objects having or not certain properties or standing or not in certain relations.

Despite the obvious differences between all three approaches, there is one thing that they all seem to have in common: they all, implicitly or explicitly, hold to a similar view of the relation between language and the world.

> One way to *misunderstand* our arguments is to take them as concerning *words but not things*. [...] It is true that we have shown that, in a relevant manner, terms for ordinary things are incoherent. In that that is so, those terms cannot apply to anything real. And from that it follows that there are no such ordinary things as those words might purport to designate. [Unger, 1979, p. 147, his emphasis]

The general idea is that words are involved in a determinate relation with the world by *designating* things. Unger and the mereological nihilists rely on this role of language to derive their negative conclusions about ontology; the defenders of ontic vagueness project the vagueness in language directly onto the world, which is only tenable by holding to a determinate relation between one and the other; and the epistemic view explicitly declares that it is partially motivated by the desire to preserve such a picture of language. But is this underlying picture a correct one? Are we actually warranted to draw such conclusions about the world based on this view?

2.4 Conclusions

Vagueness is, it appears, a very pervasive and fundamental phenomenon. It raises important questions regarding the coherence of natural language. The basic features of the problem seem to also extend to metaphysical concerns about identity and ontology. In order to appreciate the latter, however, it is crucial to understand the relation between language and the world. In the following chapter we will explore different theories of meaning and what the implications of vagueness for these theories are.

Chapter 3

Vagueness and theories of meaning

In the previous chapter, motivated by the ontological interpretations of vagueness, we started to question the assumption of a direct relation between language and the world. In this chapter, we will identify a related assumption that pervades most current theories of meaning, namely that of meaning as a form of determinate correspondence. We will discuss how vagueness can be seen as a challenge to this assumption and argue for the need to abandon it in order to more adequately understand natural language.

3.1 Meaning, correspondence, and truth

Historically, most theories of meaning in philosophy of language characterize meaning as a relation of *correspondence* between words, or more generally signs, and entities which are independent of those signs.[1] Signs, thus, are said to *have* meanings, and the role of a theory of meaning is to explicate the nature of this relation between signs and the entities they correspond to. This relation of correspondence can be equated with the notion of *reference*, which is typically construed as direct or mediated. In the first case, the meaning of a word is simply constituted by the entity or entities it corresponds to, *i.e.* the *referent* of the sign; in the second, the entities are still fundamental, but there is more to meaning than just them. The relation between the referent and the sign is said to be mediated by another type of entity or a process, which also plays a role in what constitutes meaning.

The conception of meaning as direct reference is perhaps the most common form of a pre-theoretical understanding of language. It is also to be found in many theories of meaning in philosophy of language. John Locke [1690] held the view that meaning is a direct relation between signs and the contents of consciousness. An example of a modern version of that position is implicit in the Language of Thought hypothesis championed by Jerry Fodor [1975]. There are also those who conceptualize meaning as a direct relation between signs and the world. The work of Bertrand Russell [1905] and the early philosophy of Ludwig Wittgenstein [1921] are classic examples. These ideas can still be found in modern theories of meaning like, for example, the notion of proper names as 'rigid designators' by Saul Kripke [1980].

Mediated reference goes back to the ideas of Gottlob Frege [1892]. Motivated by an identity puzzle, he defended that, although signs in general pick out referents, there must be more to meaning than just the referents. Take the expressions $a = a$ and $a = b$. Even when a and b have the same referent, according to Frege the expressions clearly have a different cognitive significance. In order to account for this, he argues that signs have not only referents, but also senses, which he describes as the 'mode of presentation' of the referents. An example Frege gives is of three lines that intersect at the same point: we can refer to the point as the intersection by means of any two

[1] It is rather difficult to try to summarize centuries of philosophical thinking about the nature of meaning in just this small section. We hope that we at least did not grossly misrepresent the essential ideas that are relevant for our discussion. For a recent and more extensive overview see, for example, Kamp and Stokhof [2008].

of the lines, thus having different possible designations of the same referent. It is important to note that, even though the correspondence between signs and entities is mediated, it is nevertheless a determinate one:

> The regular connection between a sign, its sense and its nominatum is such that there corresponds a definite sense to the sign and to this sense there corresponds again a definite nominatum; [Frege, 1892, p. 218 (reprint)][2]

This is an important characteristic that is common to both direct and mediated theories of meaning as correspondence, since it motivates the formal analysis of meaning in terms of logic.

If meaning is reference, direct or mediated, and reference is determinate, then there is a fact of the matter regarding which entities correspond to which signs. Therefore, we can analyze sentences as logical propositions and study their meaning in terms of what makes them true or false. This general approach forms the basis of most formal theories of meaning, ranging from Frege and Russell to the modern theories typically grouped under the name of *formal semantics*. A paradigmatic example is the system of logical grammar developed by Richard Montague [1970; 1970] which still forms the backbone of most modern developments in the field. One important characteristic of modern formal semantics is that it abstracts away from the nature of the entities that constitute the meanings. What is studied is not if referents are mental concepts or real objects, but purely how can words relate to entities, whatever those might be like.[3]

Despite the nuances regarding the nature of referents and the particular mechanisms involved in the reference relation, the common thread between all these philosophical positions on meaning can be summarized by the following assumptions:

1. The meaning of signs is constituted by their correspondence to certain entities, whatever the nature of those might be;

2. The use of signs in language is something essentially regular and we can study it by unveiling the rules that underlie those uses.

Theories of meaning that rely on these two assumptions we shall henceforth call *correspondence theories of meaning*. What problems arise from such theories when we consider the phenomenon of vagueness?

Let us go back to the example of the heap. Firstly, we are asked to consider a certain, large enough, number of grains of sand such that we would be ready to call it a 'heap'. In doing so, in terms of the assumptions above, we acknowledge that the entity we just considered belongs to the group of entities to which the word 'heap' corresponds. Secondly, it it sometimes explicit at this stage of the formulation of the paradox that we acknowledge that we would *not* be prepared to call one grain of sand a 'heap', thus excluding one grain of sand from the group of entities to which the word potentially corresponds. If not explicit, this realization typically comes when the paradox reaches its conclusion. Thirdly, we are asked to acknowledge a certain level of tolerance in the application of the word 'heap', by having to accept that if we would be ready to call something a 'heap', one grain of sand more or less would not make a difference towards this decision. Finally, we are confronted with the observation that we can thus repeatedly remove grains of sand from our entity originally considered and, by our tolerance principle just stated, no removal of one grain of sand would make a difference towards we calling it a 'heap'; ultimately we are lead to the conclusion that we should actually be prepared to call one grain of sand a 'heap', contrary to our initial commitment.

We can see the force of this argument as an attack on the assumptions underlying correspondence theories of meaning by revisiting the ontological interpretations of the sorites paradox.

[2]In some other translations of Frege, the german word 'Bedeutung' is translated as 'reference' rather than 'nominatum'. It is thus important to clarify that our use of 'reference' in this discussion regards the whole correspondence from sign to 'Bedeutung', rather than just the 'Bedeutung'.

[3]A great deal of work in formal semantics focuses on the compositionality of meaning, *i.e.* how complex expressions get their meaning from their individual constituents. We are not concerned with that here, but only with how the meaning of those individual constituents are typically represented as correspondence relations.

3.2 Ontology and correspondence theories of meaning

As we observed in section 2.3, most ontological interpretations of vagueness seem to rely on a conception of meaning that is in fact a correspondence theory as we have here described. As such, their considerations can perhaps be seen first and foremost as challenges to this specific picture of meaning, rather than necessarily pertaining to the actual world.

Mereological nihilism contends that vagueness leads us to conclude that there are no ordinary things, because our words for ordinary things *fail to refer* in a determinate and coherent manner. If we hold to the correspondence theory of meaning, the implication is thus that our vague words do not have any meaning. Rather than concluding that, we should perhaps see the argument as a demonstration of the failure of the first assumption of the correspondence theory of meaning. The epistemic view raises the problem of our limitations in the use of language to speak about the world. Why should we assume that we *can* know all the relevant facts that make something a heap or not? Although one of the motivations for this line of argument is to preserve classical logics and semantics, we could actually see it as undermining the relevance of the correspondence theory of meaning as a useful approach. If we cannot determinately know the referent of a vague term or expression, why should determinate reference play a role in our theories of meaning? A similar argument can be made about the belief in ontic vagueness.

We believe the phenomenon of vagueness calls indeed for a different conception of meaning alternative to the correspondence theories. Other authors have made a similar point, among them Crispin Wright:

> The picture evoked [. . .] is that there is, for any particular expression in the language, a set of such rules *completely* determinant of when the expression is used correctly; such a set thus provides a model of the information of which a master of the use of the expression may be deemed to be in possession. Clearly, however, the feasibility of such a picture requires that the rules associated with an expression, about whose use we generally agree, be consistent. For if they issue conflicting verdicts upon the correctness of a particular application of the expression, it cannot be explained just by appeal to the rules why we agree that the application is e.g. correct. [1976, p. 171 (reprint)]

One way to revise our aforementioned assumptions is to reconsider the nature of the rules that supposedly underlie our language use and how these rules can be combined. This idea lead many authors to either reinterpret classical semantics or to develop non-classical logics that incorporate a certain degree of tolerance. In the following we will briefly discuss two popular kinds of such approaches, namely supervaluationism and many-valued logics.

3.3 Supervaluationism

The position known as *supervaluationism* has been proposed and developed by several authors, the earliest accounts being those by Henryk Mehlberg [1958] and Kit Fine [1975]. The main idea behind it is that vagueness is simply a result of semantic indecision: vague predicates fail to determine precise boundaries for their application because we have not yet bothered to define them. However, we could hypothetically do so in a precise way. For example, the word 'heap' might be defined to be applicable only to piles of sand with more than, say, 100 grains. This would be one of many possible ways of making the word 'heap' precise (called a 'precisification'). The supervaluationist proposal is then to reinterpret the notion of truth so that "a vague sentence is true if and only if it is true for all ways possible of making it completely precise" [Fine, 1975, p. 119 (reprint)]. Thus, if x is a pile of sand with 101 grains, it is neither true nor false that 'x is a heap', since we can define a precisification at 100 grains, such that the sentence is true, but it would be just as reasonable to define a precisification at 102 grains, in which case the sentence would be false. However, for a pile of sand y composed of only 1 grain, there is no reasonable precisification that would make 'y is a heap' true, so the sentence is false on all precisifications, hence false simpliciter. Despite the non-classical semantics, supervaluationism claims to retain classical logic, although this is sometimes disputed [Keefe and Smith, 1999, pp. 29–32].

Without going into the details of the theory, let us reflect on its adequacy as a theory of meaning at a conceptual level. First of all, what is involved in making a sentence completely precise? Consider our example of the heap above. We gave an example of defining a precisification at 100 grains of sand. But is this *completely* precise? Clearly, a sorites sequence in terms of number of grains of sand does not affect our precisification of 'heap' anymore, since there is a specific number of grains above which something can be called a 'heap' and below which it cannot; we made sure of that by establishing a boundary. However, couldn't we conceive a different sorites sequence that would actually do the job? Think of removing atoms from our grains of sand, one by one. Would there be a precise boundary for the number of atoms that constitute a grain of sand such that we could say we have one grain of sand less? It seems rather unintuitive to say so[4]. This implies that our precisification is also vague and we should probably have specified it down to the atomic level[5].

Now let us take a step back and reflect on what this implies for supervaluationism as a theory of meaning. Supervaluationism is also a correspondence theory of meaning, albeit deviating in interpretation from classical semantics. As such, in order to determine the meaning of the sentence 'x is a heap' we are supposed to have to consider all the myriad possible precisifications, each specified down to the atomic level, and evaluate their truth value. In case the sentence is in fact true[6], this presents an unrealistic requirement for any human being to comply with. This, we believe, substantially undermines the plausibility of supervaluationism as a theory of meaning in natural language.

3.4 Many-valued logics

Many-valued logics expand the possible range of truth values to more than simply true or false. These approaches typically either add a third truth value between true and false or consider a continuous infinite range of truth values between true and false, *i.e.* degrees of truth. A theory that makes use of three truth values has to avoid a trivial pitfall: naively adding an extra truth value could result in having two problematic strict boundaries rather than one. For this reason, degree theories are typically more popular than three-valued theories.

The standard approach of degree theories is to represent truth in terms of the real unit interval $[0, 1]$, with 0 corresponding to complete falsity and 1 to complete truth. From there on, approaches differ regarding whether connectives are truth-functional or not, regarding the concrete semantics of connectives and referential terms, what the notion of validity should be, and other aspects (see Keefe and Smith [1999, pp. 35-49]). For example, one of the first fully worked out degree theory of meaning was that by Kenton Machina [1976], which uses truth-functional connectives with semantic definition as proposed by Łukasiewicz and Tarski [1956] and fuzzy set theory [Zadeh, 1965] for the semantic interpretation of referential terms. Furthermore, Machina defines validity as truth preservation in the sense that "a form of argument is truth-preserving iff its conclusions must be at least as true as its falsest premise" [1976, p. 197 (reprint)]. Using this definition, the sorites paradox comes out invalid even though all its premisses are almost completely true, which, it is argued, should suffice to explain our willingness to accept them.

Although degree theories are usually more popular than three-valued logics for purportedly avoiding strict boundaries, there are three-valued logics that acknowledge and try to solve the problem within the three-valued framework. Michael Tye [1994] presents a semantic theory that purports to do this while at the same time being invulnerable to the sorites paradox. The theory works with three truth values—truth, false, and indefinite—although the third value is "strictly speaking, not a truth-value at all but rather a truth-value gap" [1994, p. 281 (reprint)]. In order to

[4]See also Unger's [1979] different variations on the sorites paradox.

[5]Actually, according to our current theories of particle physics, namely due to quantum indeterminacy, not even that is possible, but let us just hypothetically assume that it is.

[6]Otherwise we could quickly decide based on a counterexample.

avoid strict boundaries, Tye makes use of so-called 'genuinely vague' sets: sets to which elements can either belong, not belong, or there can be no fact of the matter whether they belong or not. Furthermore, it is objectively indeterminate whether there are entities that do not fit either of these three categories. As such, vagueness exists not only at the level of the language itself, but also at the level of the meta-language that describes its truth-conditions.

3.5 Is the paradox solved?

Whether Tye's approach is successful in avoiding strict boundaries is not completely clear. Terence Horgan [1994], although criticizing most semantic theories, including degree theories, for typically portraying vagueness as less robust than it actually is, acknowledges Tye's vague sets as a fundamentally different approach. However, as he points out, a drawback of the approach is that "it obscures the underlying logical incoherence of vagueness rather than acknowledging it" [1994, p. 186 (note 25)]. This, we believe, aligns with our previous point regarding the usefulness of correspondence theories of meaning in the presence of vagueness.

The issue here—and this applies to all variants of the correspondence theory of meaning we discussed so far, including supervaluationism, three-valued logics and degree theories—is that the sorites paradox *is compelling* to natural language speakers. Most people will acknowledge that the sorites paradox drives them into a genuine contradiction. To devise a theory of meaning that is invulnerable to sorites paradoxes is to deny that such a contradiction actually arises, and thus to *misrepresent* natural language.

We believe the paradox *should not* be solved. A theory of meaning that takes vagueness seriously into account must instead explain how we can accept the premisses of the sorites paradox and the way these premisses are combined to reach its conclusion, acknowledge its conclusion as contradictory, and yet continue to use natural language in a sufficiently successful manner. Yet to diffuse the sorites paradox is an absolute necessity of any theory of meaning based on determinate correspondence. Rather than forcing natural language into the procrustean bed of our current theories, what we need is a paradigm shift that enables us to study meaning as it is, rather than as it should be.

3.6 Conclusions

In this chapter we have discussed how vagueness is problematic for correspondence theories of meaning. In particular, the conception of meaning as a determinate relation between signs and entities seems to be problematic or at least unrealistic when it comes to vague words, especially within classical logics and semantics. Alternative approaches that propose nuances to correspondence theories seem to take us further away from our intuitions about the significance of vagueness and the sorites paradox. What we need is thus a radically different approach to meaning, one that does away with the notion of correspondence as a fundamental ingredient, while allowing us to study vagueness in a manner that is in line with our intuitions. In the remainder of this thesis we will explore such an alternative in the notion of meaning as an emergent property of a complex system.

Chapter 4

The emergence of meaning

In this chapter we will temporarily leave aside the problem of vagueness in order to first reflect on an alternative to correspondence theories of meaning. In chapter 5, we will come back to it within the perspective argued for in this chapter.

4.1 Meaning and use

Ludwig Wittgenstein, originally a strong proponent of a theory of meaning as correspondence [1921], became in his later years perhaps the most well-known opponent of this idea. It is beyond the scope of this thesis to discuss all the different arguments brought forward by Wittgenstein's later philosophy. We will here focus on his alternative view on meaning, which can be summarized by the following quote:

> For a *large* class of cases of the employment of the word "meaning" – though not for *all* – this word can be explained in this way: the meaning of a word is its use in the language. And the *meaning* of a name is sometimes explained by pointing to its *bearer*. [Wittgenstein, 1953, §43, p. 25 (translation)]

Wittgenstein defended the view that the meaning of a concept is not necessarily tied to referents in the actual world, but can be better understood by looking at language use. The confusion of coupling reference and meaning arises because of the way concepts are often (at least on a basic level of first language acquisition) explained by ostension.

This idea is further explained by means of toy examples of primitive linguistic interactions. The first such example is what we might call the builders' language [1953, §2]. Two individuals, builder A and assistant B, are working with building stones (blocks, pillars, slabs, and beams). Builder A calls a building stone by name, and B retrieves that building stone for him. Contrast this *language-game* with how a child might learn how to behave in such situations. You could imagine a teacher calling a building stone by name and the learner having to point at the stone, or simply even repeat the name (to learn how to pronounce it, for example). In the original builder language, we would be inclined to say that the word 'Slab' means something like 'Bring me a slab', whereas in the learning example it means something like 'Show me the slab' or 'Repeat the word *Slab* for me'. Now suppose all slabs are broken and builder A says the word 'Slab' to B. Does it have a meaning? Certainly A knows what to expect from B in that situation. B would, however, not be able to perform his usual action of retrieving a slab (at least an intact one), but he could for example shake his head to indicate that there are no intact slabs. The word 'Slab' still has a meaning since both parties know how to use it, even though the presumed referent of the word no longer exists. What is important then, when it comes to meaning, are the actions tied to the use of the word.

Martin Stokhof summarizes the implications of Wittgenstein's later philosophy of meaning as follows:

> If we follow the close association between meaning and use that Wittgenstein's work suggests, we [cannot] but conclude that some aspects of meaning reside in the individual whereas others are determined by the community (or communities) to which the individual belongs; that there are aspects of meaning that are closely connected with mental content in the narrow sense, whereas others are intrinsically related to facts about the external environment; that there [are] biological and psychological determinants of meaning, but also defining influences from the socio-cultural environment. [Stokhof, 2013, p. 224]

These observations, we believe, can be understood in terms of an old notion that goes back to Aristotle, namely that of *emergence*. Meaning is not a property of a word, an individual, a community, or an entity in the world; it is an emergent property of a complex system as a whole and, as such, depends on all those aspects that constitute it, but is irreducible to any one of them. This is in line with Stokhof's view that "acknowledging heterogeneity also means foregoing attempts to postulate one particular aspect as somehow basic or fundamental and trying to explain all others as epiphenomena" [2013, p. 224]. In order to study meaning we cannot therefore merely focus on a word and an entity and assume a fixed determinate correspondence between them; we need to study the system as a whole.

Does this holistic view of meaning barre us from a formal study of semantics? Most formal approaches of modeling meaning that are currently popular rely on the reduction of meaning to a correspondence relation, be it direct or mediated. Are there other approaches to meaning that can formally capture our philosophical intuition of meaning as an emergent property of a complex system? We believe such a theory can be found in the framework of *signaling games*.

4.2 Signaling games

David Lewis [1969] pioneered an approach to modeling meaning based on game theory that has been recently gaining momentum in the research community, and has already helped researchers to gain original insight into different aspects of natural language[1]. Rather than providing formal definitions, we will illustrate the use of signaling games by showing how one could formalize Wittgenstein's builder language described in the previous section.

Let us simplify matters a bit: let us imagine there are only two types of building stones, slabs and pillars. Furthermore, let us say builder A can be in two states, either t_s, wanting a slab, or t_p, wanting a pillar, and the assistant B can perform two actions, either a_s, giving the builder a slab, or a_p, giving the builder a pillar. Naturally, if A is in state t_s he wants B to perform action a_s, if he's in state t_p he wants B to perform action a_p. Let us also assume that B wants to cooperate with A by doing what he expects. In game-theoretical fashion one would represent this situation by using a state-act matrix:

	a_s	a_p
t_s	$1, 1$	$0, 0$
t_p	$0, 0$	$1, 1$

In the rows we have A's possible states, in the columns we have B's possible actions. A payoff x, y in each cell corresponds to the payoff A and B receive (respectively), depending on the row and column that determine it, *i.e.* if A is in state t_s and B chooses to perform action a_s, then both players receive a payoff of 1 each. The precise numbers here are not too important; what is important is to keep in mind that each player wants to maximize their payoff, thus they represent the players preferences.

This is not yet a linguistic interaction. How could A and B align their interests in the situation described above? Obviously A cannot guess what B is going to do and B cannot read A's mind. However, we can introduce the possibility of A sending a message to B. To avoid assumptions of any pregiven meaning induced by our knowledge of the words 'Slab' and 'Pillar', let us simply talk

[1] See a recent overview by Brian Skyrms [2010].

about two possible messages, m_1 and m_2. Furthermore, let us assume the state A is in is chosen uniformly at random from the two possible states. Finally, the possible strategies for the players are: which message to send in each state, for A; and which action to perform for each message, for B. This can be represented in payoff matrix form as follows:

	$m_1 \mapsto a_s$, $m_2 \mapsto a_s$	$m_1 \mapsto a_s$, $m_2 \mapsto a_p$	$m_1 \mapsto a_p$, $m_2 \mapsto a_s$	$m_1 \mapsto a_p$, $m_2 \mapsto a_p$
$t_s \mapsto m_1, t_p \mapsto m_1$	$1/2, 1/2$	$1/2, 1/2$	$1/2, 1/2$	$1/2, 1/2$
$t_s \mapsto m_1, t_p \mapsto m_2$	$1/2, 1/2$	$1, 1$	$0, 0$	$1/2, 1/2$
$t_s \mapsto m_2, t_p \mapsto m_1$	$1/2, 1/2$	$0, 0$	$1, 1$	$1/2, 1/2$
$t_s \mapsto m_2, t_p \mapsto m_2$	$1/2, 1/2$	$1/2, 1/2$	$1/2, 1/2$	$1/2, 1/2$

The numbers are calculated based on the state-act matrix above, and the fact that both t_s and t_p can be chosen with probability $1/2$. There are three characteristic combinations of strategies in such a game: signaling systems, where for every state A is in there is a different message he can send so that B always performs his preferred action; complete miscommunication, where B does the opposite of what would be preferrable for A given his state; and pooling combinations, where either A always sends the same message, B always performs the same action, or both.

Consider the signaling system where A sends m_1 in state t_s and m_2 in state t_p and B performs action a_s when observing m_1 and a_p when observing m_2. As said before, the messages have no pregiven meaning, but in this situation, much like in Wittgenstein's discussion of the builder language, one could be inclined to say that m_1 means 'I want a slab' or 'Give me a slab' and m_2 means 'I want a pillar' or 'Give me a pillar'. Paraphrasing Lewis, in such a situation, although we never specified the meaning of the signals, "nothing important seems to have been left unsaid, so what has been said must somehow imply that the signals have their meaning" [Lewis, 1969, p. 125]. This is, we believe, a clear statement of meaning as emergent: we never define it, and yet we see it coming out of the interaction of the parts that constitute the system.

The game here described is one of the simplest examples of signaling games. The model can be modified and expanded in a myriad of different ways: varying the number of states, actions, messages, players; varying the underlying probabilities of the states; changing the payoff values to represent different situations; using a continuous state-space rather than a discrete one; giving the players the possibility to invent new messages, rather than having a fixed repertoire; and many others.

Such games are typically analyzed in terms of their solution concepts, *i.e.* which strategies will players that are trying to maximize their payoff use. This can be understood in terms of what would rational players choose, what would players using basic reinforcement learning end up doing after repeatedly playing the game, or what would a population of players playing the game in an evolutionary setting (natural variation and differential reproduction) converge to. Therefore, different assumptions regarding the required level of rationality of the players can be incorporated, allowing not only to model the complex ways humans use language, but also to understand simpler signaling systems, like those used by bacteria, for example.

Finally, when looking at populations of players continuously playing the game, one can take into account heterogeneous populations using different strategies and study which factors influence the evolution of the system into a homogeneous population or cause stable sub-populations playing different combinations of strategies to arise. One important factor is the topology of the social network that represents how likely is it for a player in the population to meet another one.

4.3 Emergence, signaling games, and Wittgenstein

Does the framework of signaling games provide us with a theory of meaning that reflects the insights of Wittgenstein's later philosophy? Without wanting to give a definite answer, especially given Wittgenstein's remarks on the role of concrete theories in philosophy, our opinion is that it is at least a step in that direction.

Signaling games provide us with a formal way of understanding meaning as an emergent property of a complex system. The study of meaning can be conducted by looking at the system as a whole, rather than reducing meaning to a simple relation of correspondence between signs and entities. In the signaling games framework, meaning is a result of the interaction of three ingredients:

Ontological constraints Linguistic interactions are constrained by the constitution of the state space, the actions available, and the relation between them represented in terms of the utility function. These are objective constraints external to the individual that ground, but do not necessarily determine, the emergence of meaning;

Dynamic interaction Meaning emerges out of a practice: individuals interact with each other and the particular ways in which they achieve some level of coordination between their actions through the use of signals constitute what we might interpret as the meaning of these signals;

Individual idiosyncrasies As a flip side of the social aspect of signaling games, particular characteristics of the individuals involved in the interactions also play a role in what kind of conventions can potentially arise. This can be reflected in the model as different levels of rationality, different strategy update procedures, etc.

We believe that signaling games truly acknowledge the heterogeneity of meaning in natural language, as was called for in section §4.1.

This the case not only within a particular model, but also regarding the use of the framework as a tool to study natural language in all of its diversity. Wittgenstein characterized language as an amalgam of different kinds of uses:

> There are *countless* kinds; countless different kinds of use of all the things we call "signs", "words", "sentences". And this diversity is not something fixed, given once for all; but new types of language, new language-games, as we may say, come into existence, and others become obsolete and get forgotten. [1953, §23, pp. 14–15 (translation)]

Within the signaling games framework, in order to study a particular aspect, or a particular kind of use of natural language, we can create a different model to try to explore the idiosyncrasies of that particular *language-game*[2]. This, we believe, is yet another dimension in which signaling games reflect Wittgenstein's insights.

4.4 A remark on correspondence

We would like to point out that we are not arguing that the notion of correspondence needs to be completely banned from the study of meaning. In our signaling game model of Wittgenstein's builder language, the two signaling systems are optimal equilibria. What this means is that, once in that situation, no player has an incentive to change his behavior; signaling systems are stable. Consider the signaling system where the speaker uses the strategy $\{t_s \mapsto m_1, t_p \mapsto m_2\}$ and the hearer $\{m_1 \mapsto a_s, m_2 \mapsto a_p\}$. Each strategy is in fact a determinate correspondence relation, but none represents the meaning of the messages by itself. Given the stable nature of this signaling system, we could perhaps describe a correspondence between m_1 and the pair $\langle t_s, a_s \rangle$. The point we want to make is that such a correspondence relation might be a reasonable *description* of the use the players make of m_1 in such a situation, but it does not have any *explanatory* power. Focusing only on such a correspondence prevents us from understanding how it is dependent on a number of different factors that enable it to come about. Moreover, seeing this correspondence as the essence of meaning prevents us from properly understanding situations where such a correspondence is problematic, for example in the case of vagueness.

[2] This will be more evident when we use signaling games to model vagueness, in chapter 5, and then identity, in chapter 7, all within the same framework.

4.5 Conclusions

In this section we explored a conception of meaning that overcomes the narrow view of– theories based on correspondence. We argued that the connection between meaning and use defended by Ludwig Wittgenstein in his later philosophy can be understood in terms of emergence, and that this idea is reflected also in the framework of signaling games. In the following chapters we will try to show how signaling games can be used to model vagueness in a simple manner that is in line with our insights so far.

Chapter 5

A model of vague signaling

In the previous chapter we argued in favor of understanding meaning as an emergent property of a complex system, rather than in terms of determinate correspondence. Signaling games, we argued, provide a formal framework to study meaning within this new perspective. In this chapter we will go back to the problem of vagueness and see how we can use signaling games to model vague languages.

5.1 Why is language vague?

In the simple example of a signaling game that we have discussed in section §4.2, vagueness is not really taken into account. In such simple signaling games, most reasonable evolutionary processes will lead to a signaling system and stay there. In that situation, no vagueness is really evident in the use of the messages: each state will trigger the speaker to always use a particular message; on receiving that message the hearer will always perform the appropriate action that maximizes both agents' payoffs. As such, there is a determinate relation between a specific state and a specific message, and between that message and a specific action. Vagueness in such signaling games would consist of a stable situation where at least one of the players would make use of a non-degenerate mixed strategy, *i.e.* a strategy which associates a non-zero probability to at least two different choices. As argued in detail by Barton Lipman [2009], in standard signaling games vagueness is always sub-optimal and thus cannot be a property of any stable situation. Given that vagueness is dominated by precision in signaling games, but in reality a large part of our language seems to be vague, we are forced to review our models to account for it.

Motivated by Lipman's result, Michael Franke, Gerhard Jäger, and Robert van Rooij [2010] propose two models that account for vagueness in natural language in terms of *bounded rationality*. The general idea is that agents have limitations that preclude them from always making the best choice when playing a signaling game. The first model looks into the impact of having memory limitations in a learning setting where agents use rational best response. The second, more general model abandons rational best response in favor of stochastic best response as the players' decision mechanism. Choice behavior of cognitively realistic agents is subject to error and is thus best modeled in probabilistic terms rather than in absolute terms. The result is that, under certain parameter conditions, the strategies of both players are vague in equilibrium.

In this chapter we will explore another possible explanation for vagueness: perceptual limitations. Although the second model of Franke *et al.* [2010] allows for perceptual limitations as one possible explanation of the agents' bounded rationality, we explore a different way of modeling this in order to disentangle perceptual limitations from the idea of (even if bounded) rational choice. However, the two models are not strict alternatives but can actually complement each other. Our objective here is not to provide *the* reason for the pervasiveness of vagueness in natural language. Many arguments have been advanced in the literature, even in a game-theoretical understanding of language[1]. The fact is that all of the proposed mechanisms might operate simultaneously, even at different levels. What we set ourselves to do in this chapter issimply to try to demonstrate how signaling games can handle vagueness in a natural fashion, with a small number of assumptions.

5.2 Perceptual limitations

Let us consider the case of a gradable adjective where an underlying metric is relevant (think of 'tall' or 'bald' and respectively height and number of hairs). Although the underlying *conceptual space* [Gärdenfors, 2000] might be, in theory, not only infinite but also continuous, for the purpose of modeling communication between agents with perceptual limitations it is safe to assume a discrete finite subset of that as relevant [Jäger, 2007, p. 554]. When it comes to height, for example, we can safely ignore infinitely tall people and rely on some finite bound. Moreover, two people differing one micrometer in height can safely be subsumed under the same value, since such difference is not even visible to our agents, thus we should be able to rely on a finite space. This much consideration of perceptual limitations is still in line with the work of Franke *et al.* [2010].

Let us define a *perceptual space* as a metric space $S = \langle T, d\rangle$ where T is a finite set of states and $d : T \times T \to \mathbb{R}^+$ a distance metric on that set. We will here limit ourselves to Euclidean spaces, thus we will talk only about T and assume d is the standard Euclidean distance, *i.e.* $d(x, y) = \|y - x\|$. We can furthermore define a notion of similarity on T as a function $s : T \times T \to [0, 1]$ which has the following properties for all $x, y, z, w \in T$:

$$\begin{aligned} x = y \quad &\Rightarrow \quad s(x, y) = 1 \\ d(x, y) > d(x, z) \quad &\Rightarrow \quad s(x, y) < s(x, z) \\ d(x, y) = d(z, w) \quad &\Rightarrow \quad s(x, y) = s(z, w) \end{aligned}$$

i.e. similarity is maximal for identical states and it is inversely related to distance.

Let M be a set of messages. We model the use of the messages as a *similarity maximization* signaling game [Jäger, 2007; Jäger and van Rooij, 2007], where the actions available to the hearer are state selections and the utility of both speaker and hearer $U(t, t')$ is dependent on the similarity between the speaker's state t and the hearer's state selection t'. The game is thus a type-matching game, where the task of the hearer is to guess the state of the speaker but, instead of an all-or-nothing payoff structure, the players are rewarded proportionally to how close the hearer got. The long-term behavior of such games evolving according to the replicator dynamics was first studied by Gerhard Jäger [2007]. It was found that such games have evolutionary stable solutions which correspond to *Voronoi tessellations* of the conceptual space, where the sender strategy partitions the space into convex regions, and the hearer strategy interprets signals in terms of prototypes of these regions. However, the particular model advanced by Jäger focused on pure strategies, thus precluding the possibility of vagueness in the use of the messages. In order to account for vagueness we need to consider mixed strategies as well.

A speaker strategy is thus a function $\sigma : T \to \Delta(M)$, which associates to each state a probability distribution over the set of messages[2]. A hearer strategy is a function $\rho : M \to \Delta(T)$, which associates to each message a probability distribution over the state space. We will henceforth speak of a *language* within the context of a signaling game as the pair of strategies $\langle\sigma, \rho\rangle$[3]. Let $\Pr : \Delta(T)$ be a probability distribution over the perceptual space which models the likelihood of

[1] See, for example, an overview by Kees van Deemter [2009], another one by Franke *et al.* [2010] (chapter 3), and an interesting recent proposal by Cailin O'Connor [2013].

[2] The set $\Delta(X)$ is the set of all probability distributions over X, which we will here assume are given by a probability mass function of type $X \to [0, 1]$. Given a function $f : X \to \Delta(Y)$, we will here abuse notation slightly and equate f with the uncurried version of f, thus representing $(f(x))(y)$ as $f(x, y)$.

[3] This is perhaps an abuse of terminology, since what is usually considered a language in philosophy or linguistics is something of much greater complexity than the signaling games we will here explore. A pair of strategies warrants being called a language only to the extent that, for example, Wittgenstein's builders' language discussed in chapter 4 does so too. In this thesis we will nevertheless knowingly sacrifice clarity of terminology in exchange for intuitive appeal.

each state occurring in a given round of the game. We can define the expected utility of a language as:

$$\mathrm{EU}(\sigma,\rho)=\sum_{t\in T}\sum_{m\in M}\sum_{t'\in T}\Pr(t)\times\sigma(t,m)\times\rho(m,t')\times U(t,t')$$

Finally, we introduce another level of perceptual limitations imposed on the agents, namely the inability to always correctly identify states. We will model this through the use of a confusion function[4] $c:T\to\Delta(T)$ which will represent the agent's probability of, given one state, confusing it with another. In the following, to simplify matters, we will equate confusion with similarity, thus the more similar two states are, the more likely they are to be confused with each other. Since confusion is a probability function, we need however to normalize the similarity function to guarantee that $\sum_{t'\in T}c(t,t')=1$ for all $t\in T$, which means we can define confusion in terms of similarity as follows, for all $t,t'\in T$:

$$c(t,t')=\frac{s(t,t')}{\sum_{t''\in T}s(t,t'')}$$

In the presence of the possibility of confusing states in these lines, players strategies are necessarily affected by this. Think of this as perceptual noise: a sender strategy for a state t and a message m gives the probability that the sender will send message m when perceiving state t but, if he is likely to perceive t as t', this probability should be partly dependent on the probability that he sends m in state t'. Similar remarks apply to the hearer. We can calculate perturbed versions of a speaker strategy σ and a hearer strategy ρ, for all $t,t'\in T$ and $m\in M$, as follows:

$$\tilde{\sigma}(t,m)=\sum_{t'\in T}c(t,t')\times\sigma(t',m)$$
$$\tilde{\rho}(m,t)=\sum_{t'\in T}c(t,t')\times\rho(m,t')$$

We can then incorporate perceptual noise into the game by using it in the dynamics. The general approach is to calculate a speaker (respectively hearer) strategy at time instant $i+1$ in two steps: first, calculate a noise-free *pre-strategy* based on the strategy at time i, using a dynamic of choice; then perturb the pre-strategy according to the equations above. An advantage of separating the strategy and pre-strategy in this way is that we are free to use whatever standard dynamics we want. In this thesis, we will use discrete-time replicator dynamics. Let $\sigma_i:T\to\Delta(M)$ and $\rho_i:T\to\Delta(T)$ be the sender's (respectively hearer's) strategy at time instant i. We can then calculate the pre-strategies of speaker and hearer at time $i+1$ (respectively $\varsigma_{i+1}:T\to\Delta(M)$ and $\varrho_{i+1}:M\to\Delta(T)$) as follows, for $t\in T$ and $m\in M$:

$$\begin{aligned}\varsigma_{i+1}(t,m)&=\sigma_i(t,m)\times\frac{\mathrm{EU}_{\sigma_i}(\rho_i,t,m)}{\sum_{m'\in M}\sigma_i(t,m')\times\mathrm{EU}_{\sigma_i}(\rho_i,t,m')}\\ \varrho_{i+1}(m,t)&=\rho_i(m,t)\times\frac{\mathrm{EU}_{\rho_i}(\sigma_i,m,t)}{\sum_{t'\in T}\rho_i(m,t')\times\mathrm{EU}_{\rho_i}(\sigma_i,m,t')}\end{aligned}$$

where expected utilities EU for a given t and m are given by:

$$\begin{aligned}\mathrm{EU}_{\sigma}(\rho,t,m)&=\sum_{t'\in T}\rho(m,t')\times U(t,t')\\ \mathrm{EU}_{\rho}(\sigma,m,t)&=\sum_{t'\in T}\Pr(t')\times\sigma(t',m)\times U(t',t)\end{aligned}$$

Their actual strategies at time $i+1$ are simply calculated as $\sigma_{i+1}=\tilde{\varsigma}_{i+1}$ and $\rho_{i+1}=\tilde{\varrho}_{i+1}$. Given a choice of dynamic, we can actually dispense with the notion of pre-strategy by merging the two

[4] Inspired by Franke [unpublished].

steps, obtaining the following definitions:

$$\begin{aligned}
\sigma_{i+1}(t,m) &= \sum_{t'\in T} c(t,t') \times \sigma_i(t',m) \times \frac{\mathrm{EU}_{\sigma_i}(\rho_i,t',m)}{\sum_{m'\in M} \sigma_i(t',m') \times \mathrm{EU}_{\sigma_i}(\rho_i,t',m')} \\
\rho_{i+1}(m,t) &= \sum_{t'\in T} c(t,t') \times \rho_i(m,t') \times \frac{\mathrm{EU}_{\rho_i}(\sigma_i,m,t')}{\sum_{t''\in T} \rho_i(m,t'') \times \mathrm{EU}_{\rho_i}(\sigma_i,m,t'')}
\end{aligned}$$

5.3 The model at work

From here on, for matters of simplicity, our examples will be based on a finite, uniformly spaced, subset of the unit interval $T \subset [0,1[$ as perceptual space. For example, in a state space with 100 states, $t_0 = 0$, $t_1 = 0.01$, and so forth. In accordance with Jäger and van Rooij [2007], we follow Nosofsky [1986] and use as similarity metric a negative exponential of the squared distance. For $t, t' \in T$ we have thus:

$$s(t,t') = e^{-\frac{d(t,t')^2}{1/\theta^2}}$$

The parameter θ controls how fast the similarity decays with growing distance, therefore we can think of it as representing the agent's *perceptual acuity*. We ran the replicator dynamics with random initial strategies σ_0 and ρ_0 until a convergence criterion was met, namely that the total absolute change in both sender and receiver strategy did not amount to more than 0.01, *i.e.*:

$$\sum_{t\in T}\sum_{m\in M} |\sigma_{i+1}(t,m) - \sigma_i(t,m)| < 0.01 \quad \wedge \quad \sum_{m\in M}\sum_{t\in T} |\rho_{i+1}(m,t) - \rho_i(m,t)| < 0.01$$

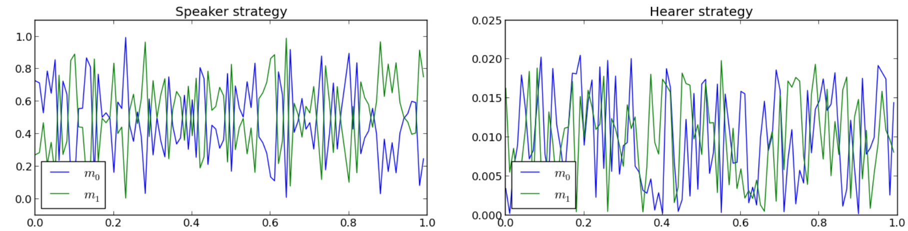
In figure 5.1 we illustrate the evolutionary process by plotting an example of the early stages of speaker and hearer strategies. In figure 5.2 we show the outcome of a number of runs with different parameter settings, as well as the progress of the expected utility of each language during the evolutionary process.

What figure 5.2 shows are vague languages. The speaker's behavior in convergence makes use of a non-degenerate mixed strategy, where for certain states the probability of a message being sent is non-zero for at least two messages. On the other hand, the hearer's behavior also makes use of a non-degenerate mixed strategy, where a message has non-zero probability of being associated with more than one state. Therefore, we believe the model captures well the behavior of agents that communicate using a vague language. Take the example in figure 5.2a. The speaker behaves very consistently when it comes to states close to the lower and upper boundaries of the state space. These are the areas where you find the determinate cases of applicability of the terms. However, there is also an area of the state space (in this case roughly between 0.4 and 0.6) where the behavior is less determined. This is the area for borderline cases. For example, think of the state space as a range of heights and the messages corresponding to the terms 'short' and 'tall'; low values of height will elicit a consistent use of 'short', high values of height will elicit a consistent use of 'tall', and average values of height will be borderline cases, not eliciting a consistent behavior from the speaker, which illustrates a level of uncertainty regarding which term to use. The hearer also exhibits uncertainty in his behavior by making use of graded prototypes.

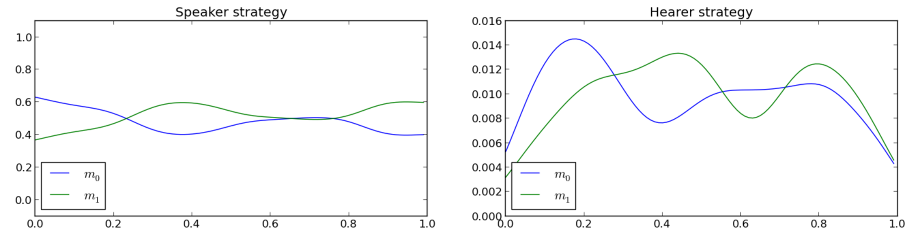
What we intend to illustrate in figure 5.2e is how the languages are beneficial to the agents, even without achieving absolute precision. During the evolutionary process, the expected utility keeps increasing; their language, even though vague, increases their level of success in coordinating their actions.

5.4 Implications for ontology

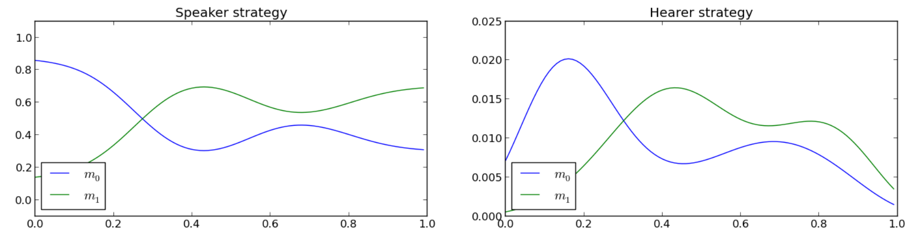
How does the approach to explaining vagueness propounded in this chapter relate to the ontological interpretations discussed in section §2.3? We believe it actually captures most insights of these



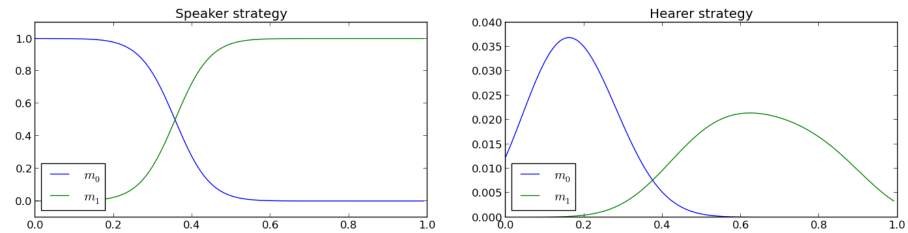
(a) $i = 0$



(b) $i = 1$

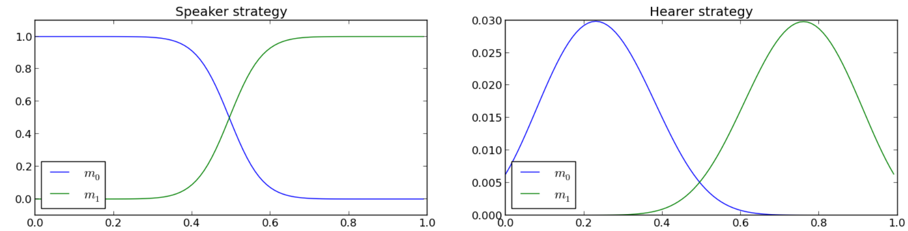


(c) $i = 3$

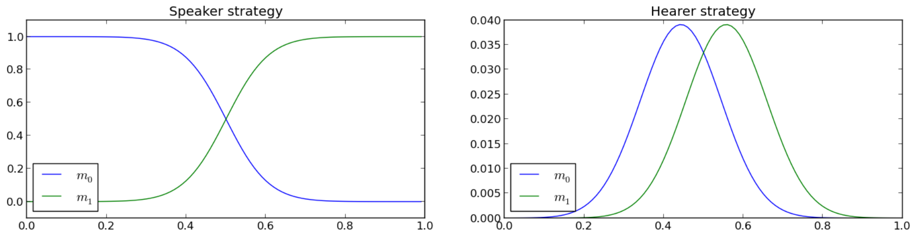


(d) $i = 7$

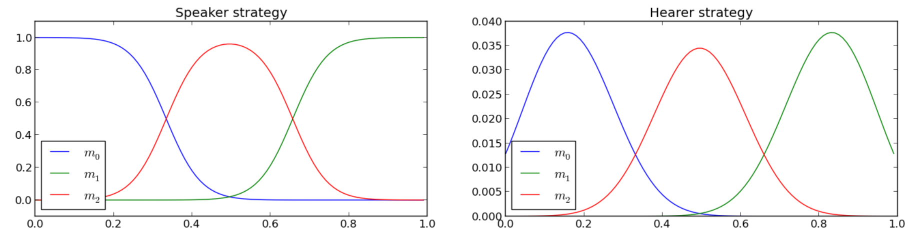
Figure 5.1: Speaker and hearer strategies during the early stages of the evolutionary process. The x axis represents the state space, the y axis probabilities. We represent the speaker strategy by plotting for each message m a function giving the probability m is used by the speaker for each state t. We represent the hearer strategy by plotting for each message m a function giving the probability a state t is selected by the hearer when receiving message m. Parameter settings are the following: 100 states, uniform priors, 2 messages, perceptual acuity $\theta = 10$.



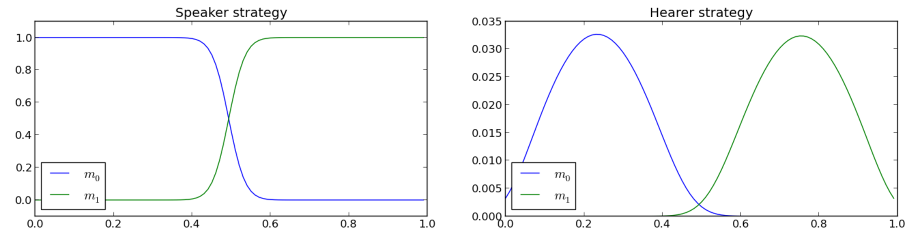
(a) 100 states, uniform priors, 2 messages, perceptual acuity $\theta = 10$.



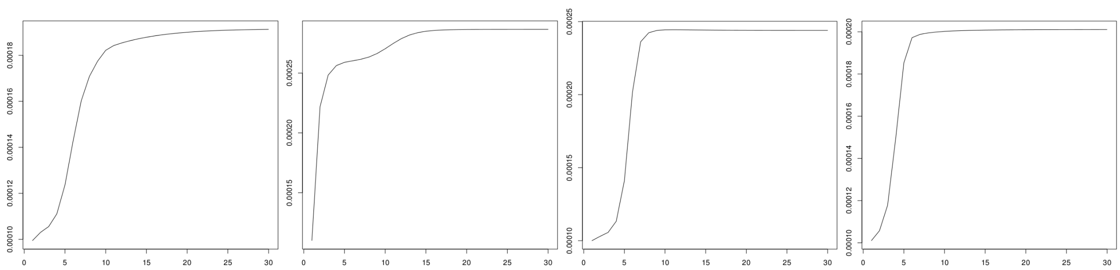
(b) 100 states, normal priors, 2 messages, perceptual acuity $\theta = 10$.



(c) 100 states, uniform priors, 3 messages, perceptual acuity $\theta = 10$.



(d) 100 states, uniform priors, 2 messages, perceptual acuity $\theta = 20$.



(e) Evolution of the expected utility of each language in the first 30 iterations of the evolutionary process.

Figure 5.2: Result of running the replicator dynamics, with some example parameter settings, until a convergence criterion was met. We also plot the progress of the expected utility of each language in the first 30 iterations of the evolutionary process.

interpretations, even though we abandoned the correspondence view of meaning, an idea central to the main lines of argumentation.

The main insight of *mereological nihilism* is that ordinary things, *i.e.* things composed of parts, do not exist. The idea is that words that purport to refer to things composed of parts are always at the mercy of a sorites paradox, thus they cannot coherently refer to the things they purport to refer, thus these things cannot exist. The modeling advanced in this chapter does away with the assumption that words necessarily purport to refer to things and explains meaning partly in terms of the behavior of agents using the words, partly in terms of evolutionary interaction processes that drive them towards coordination, and partly in terms of ontological constraints imposed on the agents by the external world. Grounded in extra dimensions of explanation, the model does not require any assumption of the existence of ordinary things. Consider the language represented in figure 5.2a. Imagine each state in the state space is a value of height and the messages correspond to the terms 'short' and 'tall'. Unlike any correspondence theory of meaning, the model explains the use of the terms without the need to rely on any set of tall people at the ontology level. In a sense, to this model the set of tall entities is a purely linguistic category and does not exist as a distinct entity, in that it does not have any special ontological status. This is very much in line with the negative conclusions of mereological nihilism, but goes even further by given a positive explanation of how use of language can be successful without these ordinary things.

The point of connection with the *epistemic view* is also quite salient. The explanation advanced for the existence of vagueness in natural language is largely based on a form of epistemic limitation of the agents involved. The extra ingredient needed to allow for vagueness in models that typically preclude it, was the assumption that there can be a non-zero probability that agents confuse one state with a different one. Thus, perceptual limitations are epistemic in the sense that such agents do not fully *know* what they are perceiving, otherwise they would not confuse one state with another. Moreover, although the connection here is perhaps more tenuous, this limitation is also ontological in the sense that it is imposed on the agents despite their level of rationality. In the model presented in section §5.1, even a best response dynamic would lead to a vague language. It is therefore a form of *ontic vagueness*, although not in terms of vague objects.

In section §2.3 we argued that all these three interpretations rely, or explicitly defend, a correspondence view of meaning. Given our earlier rejection of correspondence theories, why do we bother looking for points of connection? The reason is that we believe a lot of the intuitions behind these interpretations are very valuable, but are better understood only after we are able to move away from the idea of correspondence as a necessary construct to explain meaning.

5.5 Conclusion

In this chapter we presented a signaling game model of natural language that takes vagueness into account by enforcing perceptual limitations on the agents. The model successfully captures, we argued, the behavior of natural language users when using vague terms. Furthermore, it allows us to incorporate insights from different ontological claims motivated by vagueness, even though we abandoned the correspondence-based approach to explaining meaning. In the following chapter, we will look into how we can complete the picture by modeling the sorites paradox within this approach.

Chapter 6

The sorites paradox

In the previous chapter we saw how we can account for vagueness in a framework using signaling games, by relying on some very basic assumptions, namely a utility function based on similarity coupled with limitations in perceptual acuity. As we argued in chapter 3, without explaining the sorites paradox the picture is not complete. We should now try to understand what mechanisms underlie it, and which factors influence its success.

6.1 Modeling the argument

Let us consider an abstract formulation of the paradox. We will namely be looking at a variant often called the *forced-march* sorites paradox, where the argument is put forward step by step, rather than using a universal quantifier. Let F be a predicate and $T = \{t_0, t_1, \ldots, t_n\}$ be our state space. The two premisses involved in a sorites argument are, for $0 \leq i < n$:

$$F(t_i) \tag{6.1}$$

$$\text{if } F(t_i) \text{ then } F(t_{i+1}) \tag{6.2}$$

We will call premiss 6.1 the *base* premiss and premiss 6.2 the *tolerance* premiss. In a sorites argument, an arguer asks a speaker to assert or deny, for a certain state t_i, first the base premiss and then the tolerance premiss. One important feature of the argument that should be noted here is that, at least in the standard formulations of the paradox, the arguer elicits *bivalent* responses from the speaker. In this way, the job of premiss 6.1 is to force the speaker to commit to a decision regarding his willingness to use F in the presence of t_i; the job of premiss 6.2 is to re-acknowledge, at every step, his commitment to the tolerance principle and its consequences. The whole argument iterates these two premisses starting from a state t_i where it is not controversial that F applies and moving stepwise towards a state t_j where it is not controversial that F does not apply. The argument is successful in generating a paradox if the speaker is driven to acknowledge that F must, after all, apply to t_j. Let us see how we can model these intuitions within the framework proposed in the previous chapter.

Let $\sigma : T \to \Delta(M)$ be a speaker strategy over a state space T and a set of messages M. We start at a certain state t_i and, for that state, a message is elicited from the speaker. We will assume the speaker commits to the message that has the highest likelihood for that state, *i.e.* $m_i = \arg\max_{m \in M} \sigma(t_i, m)$. This can be represented as a pre-strategy ς_i defined as follows, for $t \in T$ and $m \in M$:

$$\varsigma_i(t, m) = \begin{cases} 1 & \text{if } t = t_i \text{ and } m = m_i \\ 0 & \text{otherwise} \end{cases}$$

The next step is to elicit an acknowledgment of the tolerance premiss. Our interpretation is that this causes the speaker to restate his commitment to the tolerance principle. We model this by perturbing the partial strategy ς_i in the same way as was done for the pre-strategy in

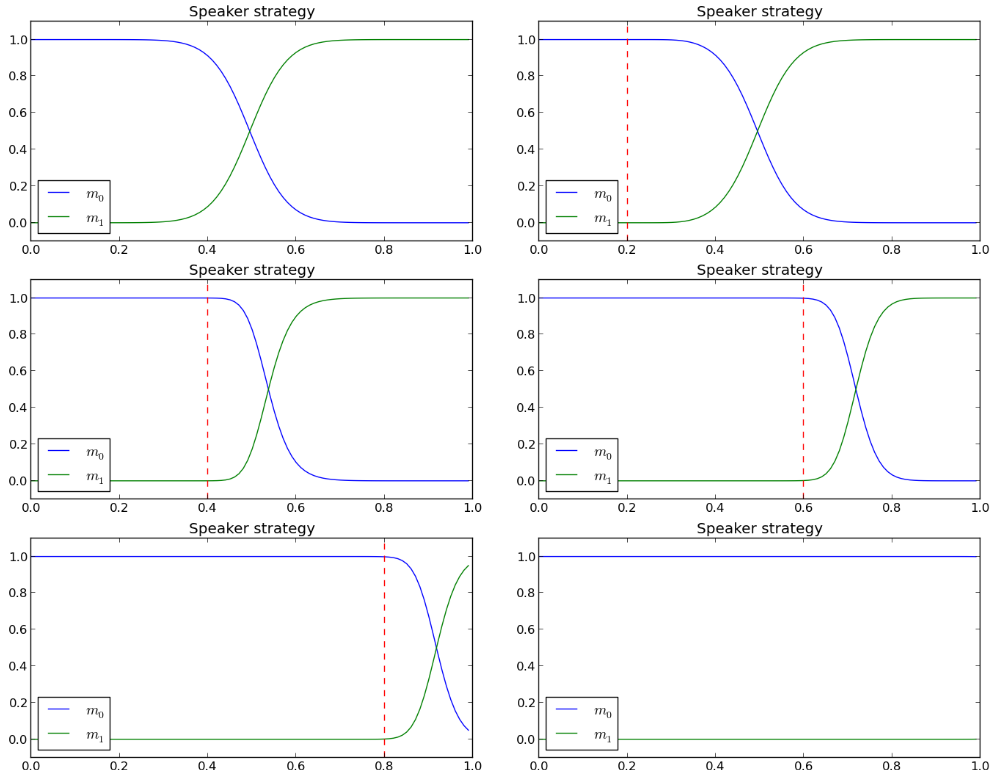


Figure 6.1: Example of a successful sorites argument going through. We plot the speaker strategy at six stages in the process. The state under focus at each stage is represented by the red vertical line.

section §5.2. The speaker's commitment to both the base premiss as the tolerance premiss should lead him to revise his speaker strategy accordingly. We model the revised strategy based on these considerations, for $t \in T$ and $m \in M$, as follows:

$$\sigma_{i+1}(t, m) = (1 - \alpha) \times \sigma_i(t, m) + \alpha \times \tilde{\varsigma}_i(t, m)$$

The parameter $\alpha \in [0, 1]$ allows for varying the susceptibility of the agent to the sorites argument, in the sense that higher values of α lead to more weight being given by the speaker to the implications of each sorites step in revising his speaker strategy.

6.2 The model at work

We exemplify this process with a simulation. We used a state space of 100 states, uniform priors, 2 messages, perceptual acuity $\theta = 10$ and susceptibility factor $\alpha = 0.5$. We ran the simulation by iterating through all the states, starting from state t_0 and following the natural order of the state space. In figure 6.1 we show the state of the speaker strategy at different stages during the process.

The outcome of this sorites argument is that one of the messages dominates the speaker strategy. Motivated to revise his strategy based on binary choices, first between two messages, then between full commitment and full rejection of the tolerance principle, the speaker has fallen into a *bivalent trap*. The reason it is a trap is that he ends up with a speaker strategy that is largely less beneficial than the one he had before (see figure 6.2), even though there were no changes in the game that justified it.

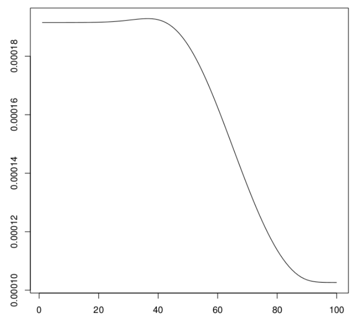


Figure 6.2: Progression of the expected utility of the language during the sorites argument plotted in figure 6.1.

In order to better understand the factors that contribute for this to happen, we performed a parameter sweep. We ran simulations with all combinations of 10 different values for number of states in the state space $|T|$, perceptual acuity θ, and susceptibility α. We formally consider a sorites argument successful if, in the final speaker strategy σ we have:

$$\exists m \in M : \forall t \in T : m = \underset{m' \in M}{\arg\max}\ \sigma(t, m')$$

i.e. there is a message that is the preferred message for all states in the speaker strategy. In figure 6.3 we plot the results in terms of the percentage of sorites arguments successful for a certain value of a parameter. To illustrate the importance of the order in the sequence of exemplars presented during the argument, we also performed the same parameter sweep but with exemplars presented in a random order, rather than in sequence.

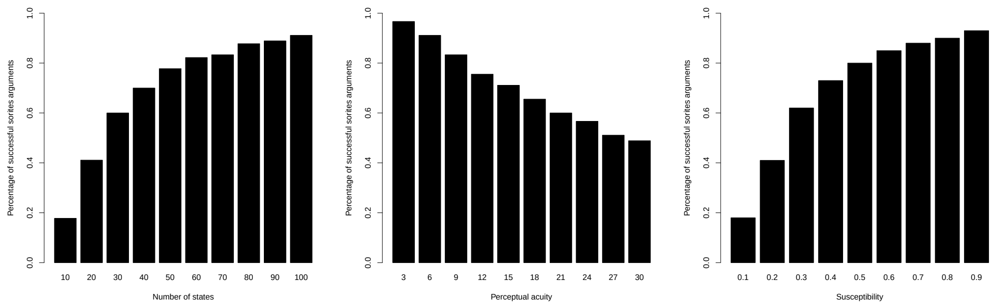
In the figure we see how the success of the sorites argument depends on a number of factors, namely the granularity of the state space, the perceptual acuity of the agents, their susceptibility to the argumentation, and finally the order in which states are put under focus. These factors can only span dimensions that we explicitly included in our models, and thus are far from an exhaustive list, but the way they affect the success of the argument can already give us some insight into the phenomenon.

One of the less insightful factors is susceptibility. Naturally we should expect that higher susceptibility leads to more likelihood of the argument succeeding; after all it controls the extent to which the speaker incorporates the outcome of each step of the argument into his existing strategy.

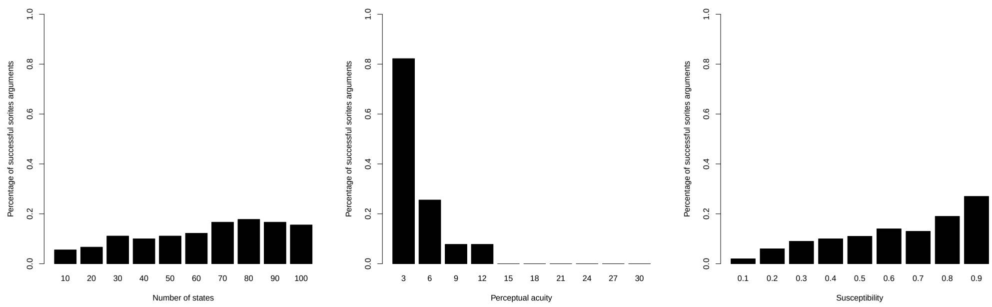
The impact of perceptual acuity should also be expected. The higher the perceptual acuity of a speaker, the easier it will be for him to distinguish states, thus the sharper the boundary will be between the use of the two terms (see again the examples in figure 5.2). This should make him less vulnerable to the paradox. However, one interesting thing to observe in the trend shown in figure 6.3a is that, even though higher perceptual acuity seems to be correlated with lower likelihood of success of a sorites argument, the values do not seem to drop considerably. This could indicate that high perceptual acuity boosts resistance to the paradox, but is not sufficient to avoid it by itself.

Regarding the order in which the argument is conducted, the contrast between the likelihood of success using a linear order (figure 6.3a) versus a random order (figure 6.3b) is quite pronounced. This shows that vagueness in language is not sufficient to drive us into contradiction; it is crucial how this vagueness is exploited in an argument. Encountering heaps of different sizes and people of different heights typically happens in a random order, rather than in an ordered sequence.

Finally, the impact of the granularity of the state space on the success of the sorites paradox can be interpreted in two ways. One is that the higher the number of smaller parts we can decompose the domain into, the more easily we can attack the use of the terms for that domain with a sorites paradox. An extreme example is the term ‘quark’, which is used for an elementary particle that is a fundamental constituent of matter. It would be ludicrous to try to argue for



(a) Sequential order argument.



(b) Random order argument.

Figure 6.3: Parameter sweep results. For each parameter we plot the percentage of successful sorites arguments per parameter value under all combinations of the remaining parameters.

the vagueness of such a term since our current theories of particle physics consider quarks as indivisible. Another interpretation of the state space granularity is in how the argument is posed. Think of the difference between arguing for the vagueness of 'tall' with a sorites argument in terms of millimeters, centimeters, and meters, and how likely would you be to succeed in each case.

6.3 Discussion

We model the sorites argument as a revision of the speaker strategy that disregards hearer strategy and utility. As such, strictly speaking it is not a game-theoretical model, although it is devised within the signaling game model of chapter 5. We effectively interpret the question 'is x a heap?' within the context of the sorites argument as 'would you call x a heap?'. The interpretation of the question in terms of the perspective of the hearer would be something along the lines of 'if I would say the word *heap*, would you think of x?'. A variant of the model could be developed from this perspective, but we believe the first one more naturally reflects the typical interpretations of the paradox. The hearer strategy can be taken into account by updating it along with the speaker according to some dynamics, for example using the replicator equations from section §5.1. The fact that the model disregards utility is a reflection of the intuition that the sorites argument questions the use of language *removed from its natural context of use*. An indicator that this is an important characteristic of the argument is its pronounced lack of success when the order of progression is not met.

Another aspect of the model we would like to discuss is how the success of a sorites argument is defined. We consider an argument to be successful in reaching a paradox whenever, after the process comes to an end, the speaker strategy is dominated by one message. This, by itself, is no paradox. The situation is only paradoxical in contrast with the speaker strategy in the beginning of the process. Take the example in figure 6.1. Whereas originally the probability of using message m_1 in state t_{100} was almost 1, in the end of the sorites argument this probability was almost 0. The assumption that the speaker is able to be aware of the paradox implies that the original strategy is somehow available for comparison, instead of having been completely replaced by the revised one. One possibility is that the agent retains the memory of the original strategy. Our preferred interpretation is that the revision of the strategy started by the sorites argument happens in parallel to the speaker's actual strategy, as a form of *hypothetical* reasoning. This is, we believe, quite reasonable since one does not start using the term 'heap' in the presence of one grain of sand, regardless of how compelling one finds the sorites argument to be.

6.4 Conclusions

In this chapter we proposed a possible model of a sorites argument within the signaling game framework proposed in chapter 5. Moving away from a correspondence view of meaning allowed us to reflect on the mechanisms behind the argument and the conditions under which a paradox can arise, all without necessarily jeopardizing our conception of language as largely beneficial to its users, and without misrepresenting the appeal of vagueness. We believe this helps to demonstrate the benefits of the paradigm shift from correspondence theories to signaling games. To further support this, we will test the generality of these considerations by addressing a different but very related problem in the next chapter.

Chapter 7

Identity and change

In section §2.2 we alluded to the close connection between the problem of vagueness and the problem of identity and change. In order to demonstrate the generality of the approach to vagueness defended in the previous chapters, we will here try to apply the same principles to explain the persistence of identity through change. We will be concerned not with so-called numerical identity but with everyday identity, *i.e.* when are we prepared to say that two things are the same? Let us revisit the Ship of Theseus paradox.

7.1 Signaling and the Ship of Theseus

There is a ship we call 'the Ship of Theseus'. In order to preserve it through the ages, whenever a plank on the ship starts to rotten, we replace it with a new one. After a considerable amount of time all planks have been replaced. Are we now still warranted to call this ship 'the Ship of Theseus' given that not even a single original plank remains? If not, when did it stop being 'the Ship of Theseus'? As with vagueness, it seems reasonable to tolerate the replacement of one plank at any stage, but the cumulative replacement of all the planks seems too much. However, we also do not want to establish a precise boundary by saying that, for example, after replacing 41 planks the ship is the same, but after 42 it is not. Can we then transfer our insights from chapter 5 to understand this problem?

It might seem odd to propose to model identity in terms of signaling games. The problem of identity is typically seen as an ontological one, rather than a linguistic one. However, there is an obvious link between the two which our formulation of the Ship of Theseus paradox tried to hint at. Namely, whether we believe that something is the same or not should be reflected in our willingness to call it 'the same'[1]. Furthermore, signaling games can be interpreted, as also argued by Jäger [2007], in a purely cognitive way. We can conceptualize states as perceptual stimuli, and the choice of a message as the choice of a more general category for a given stimulus; the choice of a category can then trigger an association with a prototypical exemplar of that category. When it comes to identity, we can think of associating a particular stimulus to some abstract representation of an object, which then in turn triggers associations of specific actions related to that abstract representation, but not necessarily the stimulus.

7.2 A model of tolerant identity

Assuming that we can thus understand identity in a manner consistent with the framework of signaling games, the connection between the model presented in chapter 5 and the Ship of Theseus paradox should be clear. Each state could correspond to a particular configuration of the ship in terms of the planks that constitute it. The notion of similarity between two states becomes

[1]We assume again, as with modeling vagueness, that we are in a fully cooperative situation.

a matter of how many planks they have in common. The original state t_0, where no plank was changed, is more similar to the state t_1, where one plank was changed, than to the state t_2 where two planks were changed. Again, the more similar two states are, the more likely the agents are to confuse them, thus perceptual limitations still apply.

However, two differences between the situations need to be addressed. First, the puzzle assumes that only one of these states actually exists at a given time. When modeling vagueness, we could assume that all states existed and there was a probability distribution that characterized how likely each state was to occur. The same approach can be taken here, but the probability distribution will be a degenerate one with all probability mass on one state, the one that is assumed to actually occur. Secondly, the model for vagueness is appropriate for two or more messages, where each of the messages ends up being associated with a prototypical exemplar within the state space. In the case of identity, given our interpretation of the state space, the conception of the problem in terms of two messages, say 'the Ship of Theseus' and 'not the Ship of Theseus', is not adequate in the same manner. The reason is that we cannot expect 'not the Ship of Theseus' to be associated with a prototype in the perceptual space, since the space consists of different potential configurations of a ship, and 'not the Ship of Theseus' could be anything.

Ideally, the second problem should be addressed by the use of only one message, but then we are faced with a technical limitation: the replicator equation used to evolve the speaker's pre-strategy ς_{i+1} will never induce any change since there are no alternative actions to sending the only message available. Therefore, we need to make two changes in the model. The first is to introduce an extra action available to the speaker corresponding to not sending any message. Because of this, the action has no counterpart for the hearer, and has no utility for any player. It is basically an option for the speaker to *opt out* of playing the game altogether. However, an action with no expected utility has no chance of being used in case the other action has non-zero expected utility, which is the case for the use of the available message if the hearer strategy associates the message to each state with a non-zero probability. The second change we need to make is to address this problem. Our solution was to introduce a cost for the speaker when the message is used.

Formally, we have a situation where the set of messages M is a singular set $\{m\}$. A speaker strategy has an extra action available, thus $\sigma : T \to \Delta(M \cup \{\perp\})$, where $\perp$ corresponds to opting out. We modify the update rule of the speaker's pre-strategy as follows for $t \in T$, $m \in M$, and $a \in M \cup \{\perp\}$, where k is a fixed cost:

$$\varsigma_{i+1}(t,a) = \sigma_i(t,a) \times \frac{\mathrm{EU}_{\sigma_i}(\rho_i,t,a) + k \times |M \cup \{\perp\}|}{\sum_{a' \in M \cup \{\perp\}} \sigma_i(t,a') \times \mathrm{EU}_{\sigma_i}(\rho_i,t,a') + k \times |M \cup \{\perp\}|}$$

The extra term $k \times |M \cup \{\perp\}|$, added to both the numerator and the denominator of the division, is necessary to avoid a negative or undefined result. Expected utilities EU are given by:

$$\begin{aligned}\mathrm{EU}_\sigma(\rho,t,\perp) &= 0\\ \mathrm{EU}_\sigma(\rho,t,m) &= \sum_{t' \in T} \rho(m,t') \times U(t,t') - k\end{aligned}$$

Everything else remains the same.

We exemplify the model by running it in similar conditions to the example situation shown in section §5.3. One of the differences is in the prior distribution, which in this example is localized at t_{50}. Another difference is that there is only one message, m_0, but the speaker has the additional opt-out strategy $\perp$. Finally, we need to choose a value for the cost parameter k, which for this example we defined as equal to the average utility, *i.e.* $k = \sum_{t \in T} \sum_{t' \in T} U(t,m,t')/|T|^2$. The resulting stable equilibrium is plotted in figure 7.1.

7.3 Modeling the Ship of Theseus paradox

The model of tolerant identity presented in the previous section results from a number of small modifications to the model of vague signaling introduced in chapter 5. The former thus retains

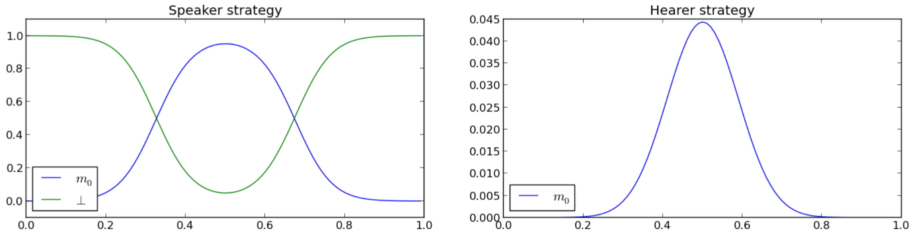


Figure 7.1: Result of running the replicator dynamics until a convergence criterion was met. Setup consisted of 100 states, degenerate priors localized at t_{50}, 1 message, an opt-out option for the speaker, perceptual acuity $\theta = 10$, and fixed message cost equal to the average utility.

many of the characteristics of the latter. Most importantly for the purpose of modeling the Ship of Theseus paradox is that it is also vulnerable to the sorites argument.

In figure 7.2 we show a sorites argument successfully going through for a case of tolerant identity with a degenerate prior distribution localized at t_0. The exact same model of chapter 6 was used without modifications. This exemplifies a possible formulation of the Ship of Theseus paradox. This particular way of modeling the paradox should be interpreted as a formulation of the problem in hypothetical terms, as argued for in section §6.3. This is so since no changes occur in the model; it is merely the speaker strategy that is revised in a manner that is removed from the conditions of actual use. As such, the fact that tolerant identity is vulnerable to paradoxes of this kind does not invalidate its utility under a natural context of use. However, there is an important difference between the sorites paradox and the Ship of Theseus.

When it comes to the sorites paradox, we argued, the conditions that enable it to arise are unlikely to occur in a natural context of use. Thus, in our natural context of use of the term 'tall' we are unlikely to be confronted with exemplars of people ordered in sequence from one that we would call 'tall' to one that we would call 'short', such that we would feel compelled to actually call the last one 'tall'. The difference when it comes to the Ship of Theseus paradox is that processes of gradual change occur all the time in our natural context of use of identity statements. Thus, even though erosion has gradually changed the Everest since its summit was first reached, we still consider it to be *the same* mountain; even though most of the cells in our bodies are 7 to 10 years old, there is a sense in which we consider ourselves to be *the same* individual that we were 15 years ago. We can model this process of natural change in our model of tolerant identity as well.

The modeling is actually very straightforward. In our model of tolerant identity, the prior distribution is degenerate, putting all probability mass on one state. This represents the state that actually occurs. In order to model natural change, the only thing we need to do is to gradually shift this prior distribution along the state space. Thus, going back to the Ship of Theseus paradox, we can start with a tolerant identity situation with prior distribution localized at t_0. After letting the evolutionary process stabilize, we can start shifting the priors in a discrete manner, by at each step i localizing the distribution at t_i. While doing this, we keep adapting the language using the replicator equations. The result can be seen in figure 7.3. What we observe is that, if we model the paradox in a natural context of use, rather than as a hypothetical exercise, identity actually accompanies the underlying change.

This helps explain why the problem is often viewed as a puzzle, rather than a paradox. In the hypothetical formulation, a paradox is reached since, we argued, the process of revision is an artificial one that occurs in parallel with the existence of a strategy adapted to a natural context of use that does not change. In this case, however, the process of revision is itself a natural one in that it occurs by continuous adaptation to actual changes in the natural context of use. That makes it more easy to defend its outcome as a reasonable conclusion, rather than a paradox. Hence, the puzzle arises: we were originally not prepared to use m_0 in the presence of t_{100}, even given our tolerance in using the message, but after the gradual change we were. In this case, we cannot even take refuge in the fact that the revision of our strategy was motivated by an artificial setting. We did not fall into a bivalent trap. What happened then?

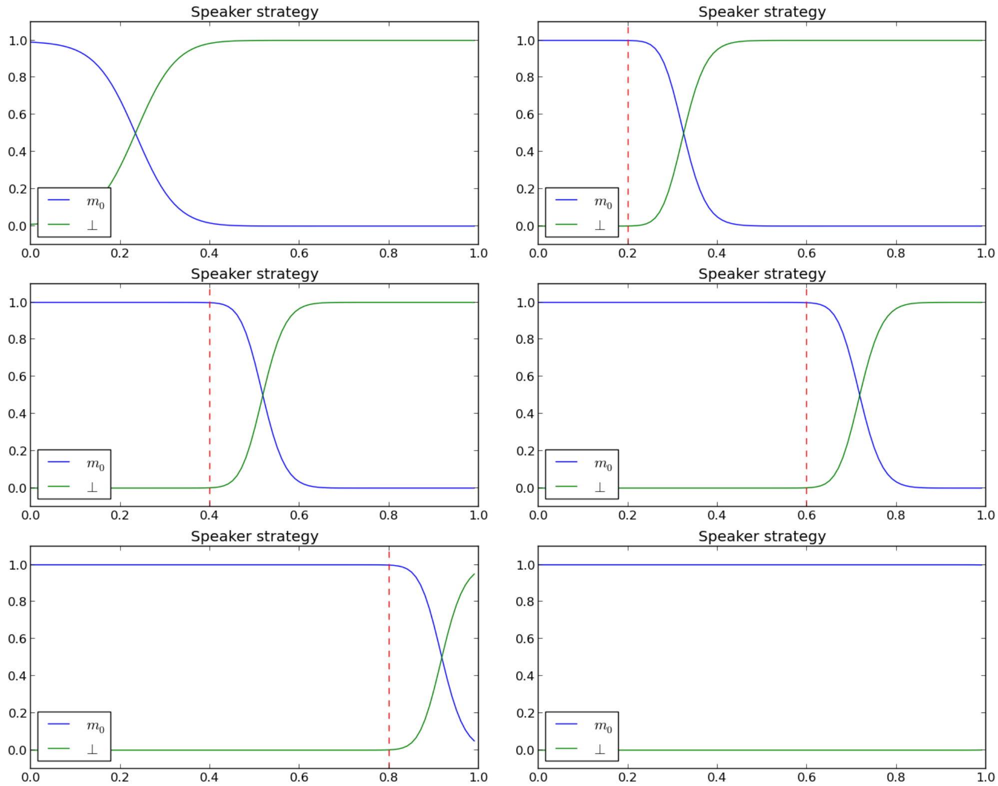


Figure 7.2: Example of a successful Ship of Theseus argument going through.

The outcome of the process of natural change is essentially puzzling, we believe, if we yet again want to rely on correspondence as an explanatory construct. Let us assume that the ship in our model was completely constituted by those 100 planks which were all replaced by the end of the process. By assuming that a term (be it 'the Ship of Theseus' or 'the same ship') has a determinate correspondence with an entity, we are puzzled that we keep using the same term even in the presence of an entity that is now completely different than the one we originally used the term for. What was then the entity the term corresponded to?

7.4 Implications for ontology

In this chapter, we tried to demonstrate how doing away with correspondence when it comes to identity statements might be the key to a more adequate understanding of the phenomenon of identity and change. By modeling identity within the framework of signaling games, we can see it also as an emergent property of a system as a whole, where correspondence plays at most a descriptive role of the agents behavior. Identity, thus seen as a pragmatic use of a signal to achieve some coordinated understanding of a state of affairs, can naturally be both tolerant and dynamic.

The ontological implications of such an approach are similar to those for vagueness discussed in section §5.4. Namely, the model advanced in this chapter does not require the existence of a distinct entity to correspond to an identity statement. The fact that we might be willing to use the expression 'the Ship of Theseus' in the presence of two distinct, but similar states of affairs does not necessarily imply the existence of a metaphysical entity that unites them both, be it an essence, a four-dimensional object, or what have you. In order to be parsimonious, we should thus avoid postulating such entities in our ontology.

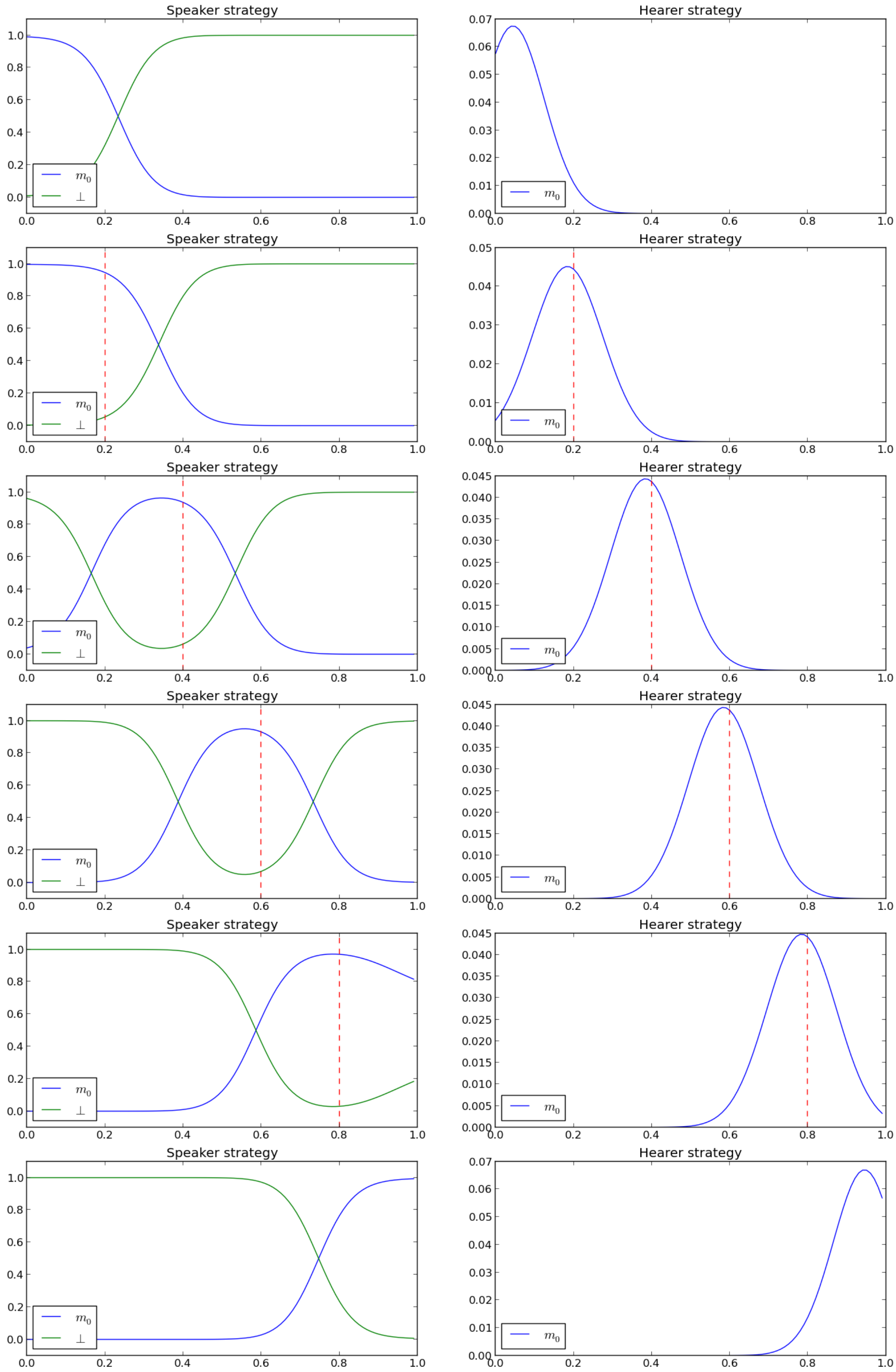


Figure 7.3: Example of identity in the presence of a process of natural change.

7.5 Conclusions

In this chapter we investigated how the insights of our approach to vagueness could be extended to a related problem: identity and change. We have shown how, with minimal changes, the model of vague signaling from chapter 5 can be transformed into a model of tolerant identity. Such model allowed us to explore how identity reacts in two different situations. We have seen how it can be vulnerable to paradoxes by arguments that exploit hypothetical change, such as the sorites argument, but nevertheless behave differently, and to the benefit of the agents, in a context of actual change.

Chapter 8

Discussion

In this chapter we would like to reflect on some of the positions defended in this thesis, discuss some related work, potential limitations of our approach, and some ideas for future work.

8.1 Language and metaphysics

In this thesis we argued for a move away from the notion of meaning as correspondence. In particular, this has implications for the relation between language and the world, and ultimately to metaphysics, as was discussed in section §3.2, section §5.4, and section §7.4. Our argument was that seeing meaning as an emergent property allows us to adequately explain many phenomena that are typically problematic for correspondence theories of meaning, such as vagueness or identity and change, without having to postulate additional entities with special ontological status. Unlike the mereological nihilists, we do not claim that such entities necessarily do not exist, but instead that there is no need to assume that they do.

Related arguments can be found in philosophy of science, for example in the debate regarding the existence of natural kinds. Conventionalists, like Nelson Goodman [1978], argue that there are many equally legitimate ways of partitioning reality in terms of kinds and thus the belief that the kinds in any of these possible conceptual schemes correspond to real entities is unwarranted. Jeffrey Barrett [2007] works out this argument in the framework of signaling games. In one of the models he explores, agents evolve languages under the pressure of an informational bottleneck, where the number of states is higher than the number of messages available. Under these conditions, and the introduction of multiple speakers, agents can develop a form of primitive grammar that is able to individuate each state through the combination of different terms. In such a situation, each term partitions the state space in the most efficient manner to allow for the individuation of all states, thus could potentially be interpreted as a natural kind. However, these partitions are purely conventional since the state space is composed of unrelated states. The upshot is that the structure of the evolved language, although beneficial for the agents, does not necessarily reflect any structure of the actual world.

These arguments, and ours, echo a form of ontological pragmatism. One example of a philosopher that has defended such an attitude is Willard Van Orman Quine:

> It is no wonder, then, that ontological controversy should tend into controversy over language. But we must not jump to the conclusion that what there is depends on words. [...] Our acceptance of an ontology is, I think, similar in principle to our acceptance of a scientific theory, say a system of physics: we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged. [...] The physical conceptual scheme simplifies our account of experience because of the way myriad scattered sense events come to be associated with single so-called objects; [...] Physical objects are postulated

> entities which round out and simplify our account of the flux of experience, just as the introduction of irrational numbers simplifies laws of arithmetic. [1948, pp. 35–37]

Our arguments in this thesis against correspondence theories of meaning reiterate the point that we should be careful when drawing metaphysical conclusions from linguistic problems, especially if we do not pay close attention to the underlying assumptions regarding the relation between language and the world. Our exploration of an alternative conception of meaning as emergent further supports this idea, by giving an account of meaning that does not require postulating the existence of physical objects. Talk of physical objects is certainly useful since, as stated by Quine, it "simplifies our account of experience". We believe, however, that if we want to gain a deeper understanding of meaning we need to go beyond these postulated entities as explanatory factors, and look instead at language from a broader perspective.

8.2 Emergence

In chapter 4 we defended a view of meaning as an emergent property of a complex system. The only way to adequately understand meaning, we argued, is to study the dynamic interplay between a variety of different elements. A similar point has been made about language in general by other authors, often with a special focus on the emergence of structure from language use. In a recent position and overview paper, the "Five Graces Group" summarizes the idea as follows:

> Recent research in the cognitive sciences has demonstrated that patterns of use strongly affect how language is acquired, is used, and changes. These processes are not independent of one another but are facets of the same *complex adaptive system* (CAS). Language as a CAS involves the following key features: The system consists of multiple agents (the speakers in the speech community) interacting with one another. The system is adaptive; that is, speakers' behavior is based on their past interactions, and current and past interactions together feed forward into future behavior. A speaker's behavior is the consequence of competing factors ranging from perceptual constraints to social motivations. The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive mechanisms. [Beckner *et al.*, 2009, pp. 1–2]

Although this description is very much in line with our discussion in chapter 4, the authors still seem to cling to a correspondence-based understanding of semantics by endorsing the view that "[t]he basic units of grammar are constructions, which are *direct form-meaning pairings*" [2009, p. 5, our emphasis]. In this thesis we thus go one step further by claiming that meaning too can be understood in terms of emergence.

To be clear, we speak of emergence in a epistemological sense, not in an ontological sense[1]. We do not claim that a genuinely new property arises at some point during the evolutionary process that leads from a random initial setup towards an equilibrium. Even in an equilibrium, meaning is not an *actual* property of the system. However, when we look at an evolutionary process as the one exemplified in figure 5.1, we are nevertheless inclined to interpret it in terms of the system moving towards a more meaningful use of the messages. Talk of meaning as emergent property is thus merely a useful way of understanding the patterns observed in the interplay between the various elements of the system.

8.3 The role of correspondence theories

A central argument in this thesis was the criticism of the notion of meaning as a form of determinate correspondence between signs and entities. Conceptually detaching meaning from the

[1] The use of the concept of emergent property is one that typically takes place in the context of philosophy of mind and philosophy of biology. See T. O'Connor and Wong [2012] for an overview of the concept and a discussion of the subtle differences across different authors, and the important distinctions one should bear in mind.

whole system that enables it promotes confusion between ontology and language, obfuscates many important factors that are crucial to understanding important phenomena such as vagueness (see chapter 5 and chapter 6), and hinders the proper appreciation of processes of dynamic adaption that, we believe, play a role in important puzzles such as identity and change (see chapter 7).

In section §4.4, however, we hinted at a possible role for correspondence theories of meaning, stressing that even though they might not have an explanatory power, they could have a useful descriptive power. Adele Abrahamsen and William Bechtel make a related point in the context of the symbolic–connectionist debate:

> the two competing approaches [...] might be given complementary roles. One way of construing this is to appreciate linguistic rules as well suited to characterizing the *phenomenon* [...] but to prefer feedforward networks as a plausible *mechanism* for producing the phenomenon. [2006, p. 167, their emphasis]

This distinction can perhaps be better understood in the context of an example from physics also discussed by the authors: the relation between pressure P, temperature T, and volume V of a gas can be described by the ideal gas law $PV = kT$; in order to provide explanations for this phenomenon we need to go down to the mechanism that is responsible for it, namely the behavior of the particles that constitute the gas, but the law itself still has descriptive, and predictive power.

We do not advocate that correspondence theories have, after all, a role in describing *meaning.* We hold to the view that meaning has to be understood in the context of a complex system. However, in a signaling games framework, correspondence theories could perhaps play a role in describing the *behavior* of the agents involved in the game. Speaker and hearer strategies like those represented in figure 5.2 can straightforwardly be described using the formal tools of degree theories, for example. It is beyond the scope of this thesis to propose concrete approaches that would best serve this purpose. However, in light of the discussions in this thesis, our opinion is that any approach that tries to do so needs to truly embrace the dynamic nature of language. Fortunately, this seems to be a direction of growing interest and development within correspondence-based theories of language[2].

8.4 Limitations of the approach

The conceptual models presented in this thesis are abstract representations that serve the purpose of argumentation. As such, they do not purport to be a realistic or complete account of natural language use. Certain simplifications are made for the sake of clarity. Here we would like to discuss to what extent these could be seen to impair the overall picture conveyed.

The first limitation we would like to discuss is the simplification of the notion of similarity. The examples used in chapter 5, chapter 6, and chapter 7 all work within one-dimensional perceptual spaces. This is not necessarily problematic if they purport to represent the use of words such as 'tall', where it does not seem controversial to consider only one dimension of comparison as relevant, namely height. We also gave the example of 'bald' as understood in terms of number of hairs, but here one could already object that the particular way the hairs are distributed might also influence the use of the word 'bald'. This becomes even more relevant for words like 'large', color terms like 'red', and ultimately identity. Conceptually this is problematic because it is not always clear how many dimensions are relevant, and which ones. However, as long as our discussion remains in the abstract, the general formulation of the model can accommodate for any finite number of dimensions[3].

Although similarity is an obvious candidate as a notion that grounds a significant part of our linguistic behavior[4], it does not paint the whole picture. Similarity maximization games might be appropriate for situations where discrimination of individual states of affairs into categories provides the agents with the highest payoffs and thus motivates their behavior. However, there

[2] See, for example, the recent work of Johan van Benthem [2011].

[3] See Jäger [2007] and Jäger and van Rooij [2007] for examples in two-dimensional spaces.

[4] See Quine [1969] on similarity and natural kinds.

are countless many ways in which we use natural language that does not fit this picture. Just restricting ourselves to gradable adjectives, we can already think of examples like the words 'nice' and 'beautiful', which not only present a challenge regarding what similarity notion would be appropriate, but also to what extent is the notion of similarity enough to characterize the motivations behind their use. We have therefore to bear in mind that the models used in this thesis are representative, at best, of particular *language-games* and do not purport to be a complete picture of our use of vague terms in natural language.

Besides trying to address these and other potential limitations in our approach, there is a number of other topics that we consider interesting to pursue in future work.

8.5 Future work

We would like to explore the connections between the model presented in chapter 5 and the work by Cailin O'Connor [2013]. The model of vague predicates she proposes is based on a dynamic of generalized reinforcement learning. The difference between this dynamic and standard reinforcement learning is that players not only "reinforce over states and acts for successful coordination, but also reinforce (to a lesser degree) on *nearby* states and acts" [2013, p. 11, her emphasis]. To what degree this reinforcement is propagated to neighboring states and acts is controlled by a gaussian function, which is a parameter in the model. Despite different phrasing and interpretation, this is very similar to the impact our confusion function has when perturbing strategies.

In order to pursue a better understanding of the phenomenon of vagueness, it would be interesting to investigate certain extensions to the model presented in chapter 5. One particular idea that comes to mind is to allow dynamic variation in perceptual acuity, the parameter which controls the level of tolerance of a player. This could allow the study of so-called linguistic hedges [Lakoff, 1973] and their role in adjusting the vagueness of other linguistic expressions. Examples of hedges are intensifiers, such as 'very', or deintensifiers, such as 'sort of' (compare the expression 'tall' with 'very tall' and 'sort of tall').

Another interesting avenue of further research would be to strengthen our results in this thesis. We studied the models in chapter 5, chapter 6, and chapter 7 in terms of simulation results; it would now be interesting to pursue analytic results in order to draw more general conclusions regarding, for example, the evolutionary outcome of the models, or the relation between the parameters of the models and the success of sorites arguments.

In the model of identity presented in chapter 7 we tried to restrict ourselves to one problem, namely identity and change. However, we believe the model can also account for other aspects of everyday identity. One such aspect is that identity is not only temporally tolerant, as is illustrated by the Ship of Theseus puzzle, but also spatially tolerant, as is demonstrated by the so-called Problem of the Many [Unger, 1980]. We would like to try to see how well our approach can handle this additional challenge.

Finally, the discussion in chapter 7 is the more exploratory work in this thesis. The objective was to investigate the potential of the change in paradigm suggested in chapter 4 by applying the same principles to a different, but related, problem. We believe that the approach is quite promising, but acknowledge that at least a proper comparison with other existing approaches should still be undertaken.

Chapter 9

Conclusions

In this thesis we explored the problem of vagueness in natural language. The problem, we argued, is extremely pervasive: most of our words seem to be vulnerable to attack by sorites arguments which reveal a conflict between how we use natural language, and how we theorize about that use. In the face of such conflict, our theories require revision. In this thesis, we focused on one particular aspect of most of our current theories of meaning that we believe should be revisited in the light of vagueness, namely the notion of determinate correspondence as an essential part of meaning.

We went on to argue for an alternative perspective: meaning as an emergent property of a complex system. In order to adequately study meaning, we argued, we cannot reduce it to a relation of correspondence but need instead to study the system as a whole. We looked into signaling games as a particular formal framework that embodies this perspective by modeling meaning as emerging from a dynamic interaction between individuals, within the context of evolutionary or cultural pressures, and under certain ontological constraints.

We proceeded to demonstrate how we can accommodate vagueness in such a model in a naturalistic way that not only allows to propose explanations of why language is vague, but– at the same time beneficial. By abandoning determinate correspondence as an essential part of meaning we were also no longer compelled to diffuse the sorites paradox. We proposed an interpretation of sorites arguments in a way that allowed us to study the conditions that facilitate its success or failure.

Finally, we tried to demonstrate the advantage of the paradigm shift by proposing a novel approach to the problem of identity and change based on our model of vagueness. We developed a notion of identity that incorporates tolerance and suggested explanations, not only of why such a notion is also vulnerable to attack by sorites arguments, but also of how it can survive a significant amount of change under natural conditions.

Bibliography

Abrahamsen, Adele and William Bechtel [2006], "Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective", in *Contemporary Debates in Cognitive Science*, ed. by Robert J. Stainton, Blackwell Publishers Ltd, chap. 10, pp. 158-185.

Barrett, Jeffrey A. [2007], "Dynamic Partitioning and the Conventionality of Kinds", *Philosophy of Science*, 74, 4, pp. 527-546.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann [2009], "Language is a complex adaptive system: Position paper", *Language Learning*, 59, December, pp. 1-26.

Cargile, James [1969], "The Sorites Paradox", *British Journal for the Philosophy of Science*, 20, 3, pp. 193-202; rpt. as "The Sorites Paradox", in Keefe *et al.* [1999], pp. 89-98.

Dummett, Michael [1975], "Wang's Paradox", *Synthese*, 30, pp. 301-24; rpt. as "Wang's Paradox", in Keefe *et al.* [1999], pp. 99-118.

Fine, Kit [1975], "Vagueness, Truth and Logic", *Synthese*, 30, 3-4, pp. 265-300; rpt. as "Truth and vagueness", in Keefe *et al.* [1999], pp. 119-150.

Fodor, Jerry A. [1975], *The Language of Thought*, Harvard University Press.

Franke, Michael [unpublished], "The Use of Quantifiers: Set Size & Typicality".

Franke, Michael, Gerhard Jäger, and Robert van Rooij [2010], "Vagueness, Signaling and Bounded Rationality", in *JSAI-isAI Workshops*, ed. by Takashi Onada, Daisuke Bekki, and Eric McCready, vol. 6797, Lecture Notes in Computer Science, Springer, pp. 45-59.

Frege, Gottlob [1892], "Über Sinn und Bedeutung", *Zeitschrift für Philosophie Und Philosophische Kritik*, 100, pp. 25-50; rpt. as "On Sense and Nominatum", in Martinich [2010], pp. 217-229.

Gärdenfors, Peter [2000], *Conceptual Spaces: The Geometry of Thought*, Bradford Books.

Goodman, Nelson [1978], *Ways of Worldmaking*, Hackett.

Horgan, Terence [1994], "Robust Vagueness and the Forced-March Sorites Paradox", *Philosophical Perspectives*, 8, pp. 159-188.

Horgan, Terence and Matjaž Potrč [2009], *Austere Realism: Contextual Semantics Meets Minimal Ontology*, Representation and Mind, MIT Press.

Jäger, Gerhard [2007], "The Evolution of Convex Categories", *Linguistics and Philosophy*, 30, 5, pp. 551-564.

Jäger, Gerhard and Robert van Rooij [2007], "Language Structure: Psychological and Social Constraints", *Synthese*, 159, 1, pp. 99-130.

Kamp, Hans and Martin Stokhof [2008], "Information in Natural Language", in *Philosophy of Information*, ed. by Pieter Adriaans and Johan van Benthem, North-Holland, pp. 49-111.

Keefe, Rosanna and Peter Smith (eds.) [1999], *Vagueness: A Reader*, MIT Press.

Kennedy, Christopher [2007], "Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives", *Linguistics and Philosophy*, 30, 1, pp. 1-45.

Kripke, Saul [1980], *Naming and Necessity*, Harvard University Press.

Lakoff, George [1973], "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts", *Journal of Philosophical Logic*, 2, 4, pp. 458-508.

Lewis, David [1969], *Convention: A Philosophical Study*, Harvard University Press.

Lipman, Barton L. [2009], "Why is language vague?", Working paper (accessed: 2013/04/02), `http://people.bu.edu/blipman/Papers/vague5.pdf`.
Locke, John [1690], *An Essay Concerning Human Understanding.*
Łukasiewicz, Jan and Alfred Tarski [1956], "Investigations into the sentential calculus", in *Logic, Semantics, Metamathematics*, Clarendon Press, pp. 38-59.
Machina, Kenton F. [1976], "Truth, Belief, and Vagueness", *Journal of Philosophical Logic*, 5, 1, pp. 47-78; rpt. as "Truth, Belief, and Vagueness", in Keefe *et al.* [1999], pp. 174-203.
Martinich, A. P. (ed.) [2010], *The Philosophy of Language*, int. 5th ed., Oxford University Press.
Mehlberg, Henryk [1958], "Truth and vagueness", in *The Reach of Science*, University of Toronto Press, pp. 256-259; rpt. as "Truth and vagueness", in Keefe *et al.* [1999], pp. 85-88.
Montague, Richard [1970a], "English as a formal language", *Linguaggi nella società nella tecnica*, ed. by Bruno Visentini, pp. 189-224.
— [1970b], "Universal Grammar", *Theoria*, 36, 3, pp. 373-398.
Nosofsky, Robert M. [1986], "Attention, similarity, and the identification-categorization relationship", *Journal of experimental psychology. General*, 115, 1, pp. 39-61.
O'Connor, Cailin [2013], "The Evolution of Vagueness", *Erkenntnis*, pp. 1-21.
O'Connor, Timothy and Hong Yu Wong [2012], "Emergent Properties", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Spring 2012, `http://plato.stanford.edu/archives/spr2012/entries/properties-emergent`.
Parikh, Rohit [1994], "Vagueness and Utility: The Semantics of Common Nouns", *Linguistics and Philosophy*, 17, 6, pp. 521-535.
Parsons, Terence and Peter Woodruff [1995], "Worldly Indeterminacy of Identity", *Proceedings of the Aristotelian Society*, 95, pp. 171-191; rpt. as "Worldly Indeterminacy of Identity", in Keefe *et al.* [1999], pp. 321-337.
Quine, Willard Van Orman [1948], "On What There Is", *Review of Metaphysics*, 2, 5, pp. 21-38.
— [1969], "Natural Kinds", in *Ontological Relativity and Other Essays*, Columbia University Press, pp. 114-38.
— [1981], "What Price Bivalence?", *Journal of Philosophy*, 78, 2, pp. 90-95.
Russell, Bertrand [1905], "On Denoting", *Mind*, 14, 56, pp. 479-493; rpt. as "On Denoting", in Martinich [2010], pp. 230-238.
Russell, Bertrand [1923], "Vagueness", *Australasian Journal of Philosophy and Psychology*, 1, pp. 84-92; rpt. as "Vagueness", in Keefe *et al.* [1999], pp. 61-68.
Skyrms, Brian [2010], *Signals*, Oxford University Press.
Sorensen, Roy A. [1988], *Blindspots*, Oxford University Press.
Stokhof, Martin [2013], "Formal Semantics and Wittgenstein: An Alternative?", *The Monist*, 96, 2, pp. 205-231.
Tye, Michael [1990], "Vague Objects", *Mind*, 99, 396, pp. 535-557.
— [1994], "Sorites Paradoxes and the Semantics of Vagueness", *Philosophical Perspectives*, 8, pp. 189-206; rpt. as "Sorites Paradoxes and the Semantics of Vagueness", in Keefe *et al.* [1999], pp. 281-293.
Unger, Peter [1979], "There Are No Ordinary Things", *Synthese*, 41, 2, pp. 117-154.
— [1980], "The Problem of the Many", *Midwest Studies in Philosophy*, 5, 1, pp. 411-468.
Van Benthem, Johan [2011], *Logical Dynamics of Information and Interaction*, Cambridge University Press.
Van Deemter, Kees [2009], "Utility and Language Generation: The Case of Vagueness", *Journal of Philosophical Logic*, 38, 6.
Van Inwagen, Peter [1988], "How to Reason About Vague Objects", *Philosophical Topics*, 16, 1, pp. 255-284.
Wheeler, Samuel C. [1975], "Reference and Vagueness", *Synthese*, 30, 3/4, pp. 367-379.
— [1979], "On That Which is Not", *Synthese*, 41, 2, pp. 155-173.
Williamson, Timothy [1992], "Vagueness and Ignorance", *Proceedings of the Aristotelian Society*, Supplementary Volume 66, pp. 145-162; rpt. as "Vagueness and Ignorance", in Keefe *et al.* [1999], pp. 265-280.
Williamson, Timothy [1994], *Vagueness*, Routledge.

Wittgenstein, Ludwig [1921], "Logisch-Philosophische Abhandlung", *Annalen der Naturphilosophie*, 14, ed. by Wilhelm Ostwald, pp. 185-262; trans. as *Tractatus Logico-Philosophicus*, trans. by C. K. Ogden, Kegan Paul, Trench and Trübner, 1922.

— [1953], *Philosophische Untersuchungen*; trans. as *Philosophical Investigations*, ed. by P. M. S. Hacker and Joachim Schulte, trans. by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte, rev. 4th ed. Wiley-Blackwell, 2009.

Wright, Crispin [1976], "Language Mastery and the Sorites Paradox", in *Truth and Meaning: Essays in Semantics*, ed. by Gareth Evans and John McDowell, Clarendon Press, pp. 223-247; rpt. as "Language Mastery and the Sorites Paradox", in Keefe *et al.* [1999], pp. 151-173.

Zadeh, Lotfi A. [1965], "Fuzzy Sets", *Information Control*, 8, pp. 338-353.

Zemach, Eddy M. [1991], "Vague Objects", *Noûs*, 25, 3, pp. 323-340.