Language, Cognition, and Mind

Richard Dietz *Editor*

Vagueness and Rationality in Language Use and Cognition

Springer

Language, Cognition, and Mind

Volume 5

This series takes the current thinking on topics in linguistics from the theoretical level to validation through empirical and experimental research. The volumes published offer insights on research that combines linguistic perspectives from recently emerging experimental semantics and pragmatics as well as experimental syntax, phonology, and cross-linguistic psycholinguistics with cognitive science perspectives on linguistics, psychology, philosophy, artificial intelligence and neuroscience, and research into the mind, using all the various technical and critical methods available. The series also publishes cross-linguistic, cross-cultural studies that focus on finding variations and universals with cognitive validity. The peer reviewed edited volumes and monographs in this series inform the reader of the advances made through empirical and experimental research in the language-related cognitive science disciplines.

More information about this series at http://www.springer.com/series/13376

Richard Dietz
Editor

Vagueness and Rationality in Language Use and Cognition

Springer

Editor
Richard Dietz
Graduate School of Arts and Sciences
The University of Tokyo
Tokyo, Japan

ISSN 2364-4109 ISSN 2364-4117 (electronic)
Language, Cognition, and Mind
ISBN 978-3-030-15930-6 ISBN 978-3-030-15931-3 (eBook)
https://doi.org/10.1007/978-3-030-15931-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The idea for this editorial project goes back to 2014 when I was invited to give a lecture series on vagueness at Seoul National University, where I met Chungmin Lee, the editor-in-chief of the *Language, Cognition, and Mind* (LCAM) series with Springer. One of the results of our discussions was Chungmin's suggestion to contribute a volume on vagueness to the then just newly launched LCAM series. I put the idea aside at first, for lack of time. A couple of years later, I came back to the idea. By that time, it had become obvious to me that the fast-growing literature on vagueness and rationality deserves more attention. If there is something like a received view about vagueness as a theoretical issue, it seems to boil down to something that may be characterised in terms of formal logic only. Undoubtedly, this logical approach to vagueness has proven very productive in the last decades. But there is reason for doubt that there is no more of theoretical interest to vagueness than a logical puzzle. For one, the orthodox model of rational decision, expected utility theory, does not supply sufficient means of accommodating vagueness in utility and in credence adequately. There is a growing sense in decision theory that there is no straightforward fix to this limitation in a way that would do justice to empirical data. Insofar as decision theory is meant to provide a predictive model of real-world behaviour, it seems fair to say that this problem is not a minor one. For another, insofar as predictive models of behaviour are guided by the methodological presumption of rationality, the question arises how to square vagueness in preferences or beliefs with constraints on rational behaviour. In particular, the question of how vague language use may be explained as a kind of rational behaviour has received increasingly attention in the social sciences. It is my hope that this volume will make this cluster of related questions more visible on the map.

I would like to thank first and foremost Chungmin Lee for suggesting me editing this volume in the LCAM series and for giving me very helpful advice in this editorial project. Work on this project was made possible through a KAKENHI Grant-in-Aid (C) for Scientific Research (No. 16K02110) awarded by the Japanese Society for the Promotion of Science (JSPS) and a Research Fellowship for Experienced Researchers awarded by the Alexander von Humboldt Foundation. My thanks go to both awarding institutions. I am grateful to Dan Lassiter and an

anonymous referee for their helpful reviews and to Nicole Standen-Mills for proofreading the whole volume. Finally, it has been a pleasure to work with Helen van der Stelt, Anita van der Linden-Rachmat and Ritu Chandwani.

Tokyo, Japan Richard Dietz

Contents

Contributors

Chrisoula Andreou Department of Philosophy, University of Utah, Salt Lake City, USA

José Pedro Correia Institute of Philosophy, University of Porto, Porto, Portugal

Richard Dietz Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

Igor Douven Sciences, Normes, Décision (CNRS), Sorbonne University, Paris, France

Michael Franke Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Matthew James Green Department of Psychology, Bournemouth University, Bournemouth, UK

Timothy W. Grinsell Holwell Shuster and Goldberg LLP, New York, NY, USA

Arnold Kochari Institute for Logic, Language and Computation, Amsterdam, The Netherlands

Anna Mahtani Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, London, UK

Nicholas J. J. Smith Department of Philosophy, The University of Sydney, Sydney, NSW, Australia

Kees van Deemter Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands;
Computing Science Department, University of Aberdeen, Scotland, UK

Robert van Rooij Institute for Logic, Language and Computation, Amsterdam, The Netherlands

Introduction

Richard Dietz

Abstract This brief introduction offers a survey of the contributed papers to this volume.

Vagueness in language and cognition has traditionally been interpreted in semantic or epistemic terms. Specifically, on the standard picture of vagueness, it is suggested that considerations of agency or rationality, broadly conceived, can be left out of the equation. Likewise, theories of rationality in choice behaviour traditionally abstract from the potential vagueness of notions that are fundamental to the theory of rational decision-making, such as credence or preference. It is in this regard fair to say that theories of vagueness and theories of rationality tend to be premised on the idea that they have nothing to contribute to one another. Recently, new literature has emerged, suggesting that this received idea is inadequate. For one, theories of vagueness in language or cognition have been used for models of rational vagueness-related credence (Edgington 1997; Dietz 2010; Williams 2012; Douven and Decock 2017; Smith 2014) and decision theory (Williams 2014).[1] For another, it has been argued that considerations of rationality should essentially factor into a more comprehensive account of vagueness. Proposals in this vain are diverse, ranging from philosophical theories on the supposed connection between vagueness and interest relativity (Fara 2000) or indeterminate projects (MacFarlane 2016), adaptations of choice theory for the semantics of vague languages (revealed preference: van Rooij 2011; social choice: Grinsell 2012), Bayesian models of pragmatic reasoning (Lassiter and Goodman 2017), or applications of evolutionary game theory to the theory of vagueness (for an overview, see Franke and Correia 2017).

This volume presents new conceptual and experimental studies that explore interconnections between vagueness in language and cognition and the theory of rational behaviour. The first three papers, by Mahtani, Smith, and Andreou, turn on the ques-

[1] The literature on imprecise credence (Bradley 2014) and imprecise preference (Hsieh 2016) is vast, but rather disconnected from the philosophical and linguistic literature on vagueness.

R. Dietz (✉)
Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
e-mail: richarddietz22@gmail.com

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*,
Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_1

tion of how vagueness affects rational choice and related fundamental notions. The other cluster of papers, by Green and van Deemter, Correia and Franke, Douven, Grinsell, and Kochari and van Rooij, bring models of rationality to bear to the theory of vagueness in language and cognition.

1. *Vagueness in Rational Choice*: Mahtani and Smith present different positions on the controversy whether rational credence may be imprecise. While Mahtani essentially agrees with previous philosophical arguments for imprecise cre-dence, she disagrees with the common way of modelling imprecise probabilities by means of sets of admissible probability functions. Mahtani argues that a supervaluationist approach to imprecise probabilities offers a more adequate means of modelling comparisons in probability between agents, or between times for a fixed single agent.

Smith distinguishes between two conceptions of credence, a dispositional conception (the disposition to act on a given proposition) and an epistemic conception (the degree to which belief in a proposition is justified by given evidence). He grants that similar to basic fuzzy theories of vagueness, any theory of epistemic credence that commits to precise probability assignments is open to the objection that this precision is artefactual. However, he argues that Bayesian models of credence in the dispositional conception, which commit likewise to precise assignments, are defensible against this kind of objection.

Vague preferences may give rise to regret regarding past opportunities that have been missed through making a choice, while at the same time there is a continued endorsement for the choice that was eventually made. A significant amount of the literature on this phenomenon focuses on the multidimensionality of preference relations as an explanation for preferential vagueness. Andreou's paper brings the acyclicity of preference relations to the fore as another possible explainer for preferential vagueness and related regret with a continued endorsement.

2. *Rationality in Vague Language Use and Cognition*: Is vagueness merely a defect in agents' ways of representing reality, in language or cognition? Or is it in some way useful for agents to represent reality in vague ways rather than in precise ways? In their experimental study on vague referential noun phrases, Green and van Deemter put the hypothesis that vagueness *is* of instrumental value to a test. The authors explain that their results are rather negative, and that observable advantages in vague language use can be explained away by other features of linguistic expressions.

In contrast to Green and van Deemter's experimental study, the conceptual studies by Correira and Franke, Douven, and Grinsell offer arguments for the idea that there is a rationale for representing reality in vague ways rather than in precise ways. Correia and Franke provide a novel argument of this kind from evolutionary game theory. In previous arguments of this type, it has been suggested that on certain provisos, best communication strategies for speakers and hearers in signalling games should be vague. The authors suggest that language communities with vague strategies, in the long run, prevail over communities with precise strategies.

Douven's account of vagueness in cognition is set out in a conceptual spaces framework, where concepts are represented as extended areas in a multidimensional space, with the dimensions being types of features relevant to categorisation. The argument is premised, first, on constraints on best solutions for categorisation problems and, second, on an empirical assumption about the features of proto- typical instances of concepts. According to this account, concepts are bound to be vague, provided that they satisfy certain elementary constraints on rational cate- gorisation and they go along with typicality judgements that are in some sense fuzzy.

Paradoxes of vagueness involve so-called sorites series, where adjacent items are indistinguishable in relevant respects (to the application of a term in question), with indistinguishability being an intransitive relation. Grinsell's basic idea is that vague terms are multidimensional, and that categorisations along relevant dimensions are choice functions, the aggregation of which can be modelled in terms of social choice theory. On this account, the sorites-susceptibility of certain terms may be explained as being due to the acyclicity of social preferences, which are due to conservative constraints on rational social choice and structural features of the relevant domain of items to be categorised.

Another way of modelling the semantics of vague terms in terms of choice functions is defended by Kochari and van Rooij. According to them, the relevant indistinguishability relation in sorites paradoxes can be understood as a relation of being different to a relevant magnitude that is lower than a given threshold. The authors furthermore present experimental results on gradable adjectives that shed new light on the relationship between vague implicit comparatives (x is red while y is not red) and associated explicit comparatives (x is redder than y).

References

Bradley, S. (2014). Imprecise probabilites. In E. Zalta (Ed), *The Stanford encyclopedia of philosophy* (Winter, 2016 Edn). https://plato.stanford.edu/entries/imprecise-probabilities/.

Dietz, R. (2010). On generalizing Kolmogorov. *Notre Dame Journal of Formal Logic, 51,* 323–335.

Douven, I., & Decock, L. (2017). What verities may be. *Mind, 126,* 386–428.

Edgington, D. (1997). Vagueness by degrees. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader* (pp. 294–316). Cambridge/MA: MIT Press.

Fara, D. G. (2000). Shifting sands: An interest relative theory of vagueness. *Philosophical Topics, 28,* 45–81.

Franke, M., & Correia, J. P. (2017). Vagueness and imprecise imitation in signalling games. *The British Journal for the Philosophy of Science, 69,* 1037–1067.

Grinsell, T. W. (2012). Avoiding predicate whiplash: Social choice theory and linguistic vagueness. *Proceedings of SALT, 22,* 424–440.

Hsieh, N. (2016). Incommensurable values. In E. Zalta (Ed), *The Stanford encyclopedia of philosophy* (Winter, 2016 Edn). https://plato.stanford.edu/entries/value-incommensurable/.

Lassiter, D., & Goodman, N. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese, 194,* 3801–3836.

MacFarlane, J. (2016). Vagueness as indecision. *Proceedings of the Aristotelian Society, 20,* 255–283.

Smith, N. J. J. (2014). Vagueness, uncertainty, and degrees of belief: Two kinds of indeterminacy—one kind of credence. *Erkenntnis, 79,* 1027–1044.

van Rooij, R. (2011). Vagueness and linguistics. In G. Ronzitti (Ed.), *Vagueness: A guide* (pp. 123–170). Berlin: Springer.

Williams, J. R. G. (2012). Generalized probabilism: Dutch books and accuracy domination. *Journal of Philosophical Logic, 41,* 811–840.

Williams, J. R. G. (2014). Decision making under indeterminacy. *Philosophers' Imprint*, *14*.

Part I
Vagueness in Rational Choice

Vagueness and Imprecise Credence

Anna Mahtani

Abstract In this paper I investigate an alternative to imprecise probabilism. Imprecise probabilism is a popular revision of orthodox Bayesianism: while the orthodox Bayesian claims that a rational agent's belief-state can be represented by a single credence function, the imprecise probabilist claims instead that a rational agent's belief-state can be represented by a set of such functions. The alternative that I put forward in this paper is to claim that the expression 'credence' is vague, and then apply the theory of supervaluationism to sentences containing this expression. This gives us a viable alternative to imprecise probabilism, and I end by comparing the two accounts. I show that supervaluationism has a simpler way of handling sentences relating the belief-states of two different people, or of the same person at two different times; that both accounts may have the resources to develop plausible decision theories; and finally that the supervaluationist can accommodate higher-order vagueness in a way that is not available to the imprecise probabilist.

1 Introduction

On the orthodox Bayesian account, every rational agent has a precise credence (or degree of belief) in every proposition that she entertains. Many have objected that this claim is implausible. A rational agent does have a precise credence in *some* propositions: for example, if I am about to toss a coin that you know to be fair, then your credence that (HEADS) it will land heads is presumably exactly 0.5. But now consider the proposition (SARDINES) that my neighbour has at least one tin of sardines in her kitchen cupboard. What is your credence in SARDINES? There are reasons to think that you don't have any precise credence in this proposition.

One reason is that nobody knows what your credence is in SARDINES—not even you. If you are asked what your credence is, then it is likely that no particular number will spring to mind. If pushed, you may be able to produce a number, but

A. Mahtani (✉)
Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, London, UK
e-mail: a.mahtani@lse.ac.uk

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*, Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_2

the number that you produce will be arbitrary. You might say, for example, ‘0.352’, but you could just as easily have said ‘0.353’. You do not *know* what your credence is in this claim. And it seems odd that you might have a precise credence—and so be in some particular mental state—without knowing what it is.[1]

Another reason is this: it is not clear why your credence in SARDINES is some particular value (0.352, say), rather than some other nearby value (such as 0.353). What is it about you that makes it the case that your credence in this proposition is exactly 0.352?[2]

Here we might think that both of these problems can be easily dealt with, for on the orthodox Bayesian view, there is a tight relationship between an agent’s credence function and her dispositional betting behaviour. Here is Bruno De Finetti on the subject:

> One can…give a direct, quantitative, numerical definition of the degree of probability attributed by a given individual to a given event…. It is a question simply of making mathematically precise the trivial and obvious idea that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event.
>
> Let us suppose that an individual is obliged to evaluate the rate *p* at which he would be ready to exchange the possession of an arbitrary sum S (positive or negative) dependent on the occurrence of a given event E, for the possession of the sum *p*S; we will say by definition that this number *p* is the measure of the degree of probability attributed by the individual considered to the event E (De Finetti 1964, pp. 101–2).

To illustrate De Finetti’s method here, we can apply it to elicit your credence in HEADS. We suppose that you are forced to bet with a bookie over HEADS, and the way the bet works is that you give the bookie some sum *p*S, and in exchange you will get the sum S if and only if HEADS is true. Before the bet is settled, you get to name the rate *p* (your ‘betting quotient’), and then the bookie gets to fix the sum S. The bookie can fix this sum as either negative or positive, and the idea is that this forces you to produce as the rate *p* your true credence in HEADS. To see why this is, suppose first that you give some high figure such as 0.8 for rate *p*. Then the bookie will set the sum S as a positive value—let’s say as £10. Then you give the bookie £8 (for this is 0.8(£10)), and the bookie will give you £10 back if and only if HEADS is true. Thus you are left committed to a bet which is—by your own lights—a bad deal for you. Now suppose instead that you give some low figure such as 0.2 for rate *p*. Then the bookie will set the sum S as a negative value—let’s say—£10. Then the bookie will give you £2 (for this is equivalent to your giving the bookie 0.2(–£10)), and you will have to give the bookie £10 if and only if HEADS is true. Again, you are left committed to a bet which is a bad deal for you. The only way to ensure a neutral deal is to set *p* equal to your own credence in HEADS—i.e. 0.5.

---

[1] At any rate, this idea seems odd at first, though an externalist about knowledge may be easily reconciled to it. Thanks to an anonymous referee for pointing this out.

[2] There may also be other reasons to doubt that you do or should have a precise credence in SARDINES. For example, James Joyce would argue that your evidence does not justify any particular credence (Joyce 2010).

Can we similarly use this method to elicit your credence in SARDINES? Presumably if we elicited your betting quotient as De Finetti recommends then you would manage to produce some rate *p*, but the number that you produce would be arbitrary. You would have no good reason to choose the number 0.352, say, over 0.353. Forced to pick a particular number, you might decide on a whim, or choose at random. We can see this clearly by thinking about the betting quotients that you would produce across close possible worlds, where you have the same evidence and rationality as in the actual world. If your betting quotient is elicited in several of these worlds, then the answers you give across these worlds will vary. For you are just deciding randomly or on a whim, and the results of these random or whimsical processes will vary across close possible worlds. We can contrast this with the case where we elicit your betting quotient for HEADS: presumably in each close possible world where you have the same evidence and rationality as you have in the actual world, you will produce the very same number (0.5) when your betting quotient for HEADS is elicited. In this way your betting behaviour across close possible worlds is *stable* where HEADS is concerned, but *unstable* where SARDINES is concerned. I have argued elsewhere (Mahtani 2016) that this sort of instability in betting behaviour is typical of the sorts of cases that motivate theorists to resist the orthodox Bayesian's claim that a rational agent has a precise credence in every proposition that she can entertain.

Thus defining an agent's credence in terms of her betting quotient has not helped. Intuitively, there is no particular number that is your credence in SARDINES, and similarly there doesn't seem to be any particular number that is your betting quotient for SARDINES—for there is no single number that is *the* number that you would produce were we to elicit your quotient as De Finetti recommends. How then should we respond? In the next section I consider (and set aside) a nihilist position, according to which you have no betting quotient, and no credence in SARDINES. Then in Sect. 3 I set out an alternative theory: that the expression 'credence' is vague.

2 Nihilism

I start by considering a nihilist position—in order to set this position aside. On this nihilist view, you do not have a credence in SARDINES. We might argue for this as follows[3]:

1. Your credence in SARDINES is the number that you would produce were your betting quotient to be elicited.

[3]This argument rests on the assumption that your credence in SARDINES is identical to your betting quotient in SARDINES, which of course is open to challenge. But it is not clear how the claim that your credence is not identical to your betting quotient can help us here: there are strong intuitive reasons to doubt that you have any particular credence in SARDINES, and the claim that your credence is identical to your betting quotient introduced in the hope that this would fix your credence in SARDINES.

2. Thus your credence in SARDINES is the number that you produce in the closet possible worlds in which your betting quotient is elicited.
3. But there are a range of equally close worlds in which your betting quotient is elicited, and the number that you produce varies across these worlds.
4. Thus there is no such thing as *the* number that you produce in the closest possible worlds in which your betting quotient is elicited.
5. Thus there is no such thing as your credence in SARDINES.

We could generalize this argument to show that you have no credence in many of the propositions that you entertain. Thus if you have a credence function at all, it does not map each proposition that you entertain onto a number between 0 and 1. Rather, at best it maps *some* of the propositions that you can entertain (such as HEADS) onto numbers between 0 and 1: your credence function does not map propositions like SARDINES to anything whatsoever. Is this a plausible position? And how can we resist the argument for it given above?

I want to start by showing that we can produce an argument paralleling that above to the conclusion that you have no resting heart rate. This seems like a surprising conclusion. Intuitively you *do* have a resting heart rate, even if you are not currently resting. Perhaps you are reading this paper while working out on a treadmill, and so your heart rate is currently elevated, but if so it would still make sense for a trainer to ask what your resting heart rate is—perhaps to check that your current training session is pitched at the right level. Thus intuitively you do have a resting heart rate, even when you are not resting. And this is because your resting heart rate is a dispositional property: it is the rate at which your heart *would* beat *were* you at rest. The problem is that there are a variety of ways to rest, and you could just as easily rest in one way as another. For example, you could rest by lying down in a cool room; or you could rest by sitting up in a warm room. Plausibly there are a range of equally close counterfactual cases where you are resting, and your heart rate will vary across these cases because your heart rate is sensitive to all sorts of factors. There are some guidelines that specify more precisely what is meant by 'resting' (Palatini 2009), but because these inevitably fall short of exact precision, your heart rate may vary even across close counterfactual cases where the guidelines are met. Thus we can construct the following argument:

1. Your resting heart rate is the rate at which your heart would beat were you at rest.
2. Thus your resting heart rate is the rate at which your heart beats in the closest possible worlds in which you are at rest.
3. But there are a range of equally close worlds at which you are at rest, and the rate at which your heart beats varies across these worlds.
4. Thus there is no such thing as *the* rate at which your heart beats in the closest possible worlds in which you are at rest.
5. Thus there is no such thing as your resting heart rate.

We could also argue along similar lines that you have no height. This is counterintuitive, for surely you do have a height—something in the region of 5′10′, say. This is your height even if you are sitting down, and thus currently measure considerably less than $5'10'$ across every dimension. If you are sitting down in the doctor's office, and she asks for your height, you do not need to check how you are currently oriented to be able to answer. For your height is a dispositional property. It is the distance that there *would* be between the top of your head and the soles of your feet *were* you to stand in a particular way: straight, but not on tiptoe. So your height depends on what this distance is in the closest worlds where you are standing in the relevant way. But the problem is that there will be a range of equally close worlds where you are standing in the relevant way: ones where you stretch your neck out slightly more; ones where you are imperceptibly slouching, and so on. The distance between the top of your head and the soles of your feet will vary across these possible worlds. Measurements of a person's height vary a surprisingly large amount: the measurement depends on many factors, including precisely how you are standing, and whether traction is applied as you are measured (Buckler 1978). Thus there is no such thing as *the* distance that there would be between the top of your head and the soles of your feet *were* you to stand up straight. Thus (following the argument pattern above) we can show that you have no height.[4]

If we accept the conclusions of these arguments, we arrive at a sort of nihilism: you have no height, no resting heart rate, and no credence in many of the propositions that you can entertain. Is this a plausible position? Well, it seems more plausible than Peter Unger's radical position on vagueness (Unger 1979). Unger has argued that there aren't any tall objects, nor bald objects, nor indeed any 'ordinary things': this is his response to the sorites paradox—a paradox that I discuss in the next section. Unger's position is very counterintuitive, because he denies many claims that we take to be uncontroversially true—such as the claim that a person who is $6'5''$ is tall. If we deny that you have a heart rate, a height, or a credence in SARDINES, do we similarly deny claims that are uncontroversially true? It is much less obvious. For though at first blush it seems obvious that you have a resting heart rate, say, this intuition is not very robust. Once we start to think about what your resting heart rate is *exactly*, it becomes clear that there is no n such that it is uncontroversial that your resting heart rate is n: thus the intuition that you have a resting heart rate is itself controversial, in contrast to a truly uncontroversial claim, such as the claim that a person who is $6'5''$ is tall.

Should we then accept that you have no resting heart rate, and no height, and so on? This would leave us with some puzzles. What is happening when the doctor asks for your resting heart rate, and you utter some number, and she takes that number into account in her assessment? If you had no resting heart rate, how could we explain

[4]With the concept of 'height' there is an added complication. We know that your height is the distance between the top of your head and the soles of your feet in some close possible world(s) in which you are standing in the right sort of way—but we do not know how to choose between the various close possible worlds that seem to meet this criterion. This is the problem described above, which has an obvious analogue for the concepts of 'resting heart rate' and 'betting quotient'. But for the concept of 'height' there is a further problem: we do not even know in any given possible world what the distance is from the top of your head to the soles of your feet. This is because we don't know where your head or feet end—which molecules of dead skin to include—for the boundaries to your body are vague.

this exchange?[5] We cannot hope to explain it as make-believe—as we might explain some dialogues involving fictional predicates—for the doctor takes your contribution seriously and uses it to guide her diagnosis. Furthermore, even though there may be no uncontroversial truths of the form *your resting heart rate is* n, there seem to be other uncontroversial truths in the vicinity. For example, it may be uncontroversially true that your resting heart rate is higher than 20 bpm, and less than 200 bpm. How can we explain the uncontroversial truth of these claims if you have no resting heart rate?

My suggestion is that we claim that you *do* have a resting heart rate, but that it is a vague matter what this resting heart rate is. The same can be said for your height, and for your credence in SARDINES. In the next section, I consider what it means to say that these predicates are vague—given that they do not fit the usual mould.

3 Vagueness

Philosophers usually introduce the phenomenon of vagueness with archetypal one-place predicates, such as '… is bald', '… is tall', '… is a small number', and so on. Vague predicates are taken to have certain features: they typically have both borderline cases and clear cases, appear to lack sharp boundaries, and are susceptible to sorites paradoxes (Keefe 2000, p. 6).

Can predicates with different forms be vague as well? Rosanna Keefe argues that they can, and that '[a] theory of vagueness should have the resources to accommodate all the different types of vague expression' (Keefe 2000, p. 14). I claim that predicates such as '… at … has a resting heart rate of…' are vague expressions—even though they do not seem to have the features typical of vague expressions. Below I work through these features in turn:

(1) *Vague terms typically have both borderline cases and clear cases*

We can use the archetypally vague predicate '… is bald' to illustrate this feature. A person with 0 hairs on his head is clearly bald; a person with 500,000 hairs on his head is clearly not bald; and a person with 5739 hairs on his head may be a borderline case. Thus this predicate has an extension with some objects clearly falling within it, some objects clearly falling outside it, and some borderline cases.

Now let us turn to our predicate '… at … has a resting heart rate (in bpm) of…' and see whether this predicate also exhibits this feature. This is a three-place predicate: if anything belongs within its extension, it will be ordered triples consisting of an animal, a time, and a number. It is easy enough to find clear cases of ordered triples that do not fall within the extension of this predicate: for example, the ordered triple

[5]We might try saying that your resting heart rate at any time is just whatever it was most recently measured as—but this is not plausible, because it makes sense to think that your resting heart rate has increased or decreased since it was last measured. And of course we might reasonably form a conjecture about a person's resting heart rate even if we knew that person had never had this measurement taken.

consisting of myself, now, and the number 20 clearly falls outside the extension of this predicate, because I certainly don't currently have a resting heart rate of 20 bpm. It is also easy enough to find examples of ordered triples that are borderline cases. For example, the ordered triple consisting of me, now, and the number 75.047 is a borderline case: it is not clearly the case that my current resting heart rate is 75.047 bpm, but it is not clearly not the case either. But it seems that we cannot find any examples of ordered triples that clearly fall within the extension of the predicate, for no animal at a time clearly has a resting heart rate of precisely *n*, for any number *n*. Thus it seems that this predicate only partially displays this feature of vague predicates.

(2) *Vague terms appear to lack sharp boundaries*

Again, let us use '… is bald' to illustrate this feature. Some people fall within the extension of this predicate, and some fall outside. We can then visualize a boundary separating those who are bald from those who are not. But because there are borderline cases, intuitively there is no *sharp* boundary between those who are bald and those who are not. We might instead imagine two boundaries, one separating out the objects that clearly fall within the extension of the predicate, and one separating out the objects that clearly fall outside the extension of the predicate, leaving the objects in between these two boundaries as the borderline cases. But intuitively these boundaries are not sharp either: for there are objects that are not clearly borderline. More generally, archetypal vague predicates seem to lack sharp boundaries altogether.

Does this also hold for predicates such as '… at … has a resting heart rate (in bpm) of…'? Here the whole concept of a boundary is harder to make sense of. Where should we look for a candidate boundary to see whether it is sharp? If we consider all the possible ordered triples of objects, then we will find that a few of these ordered triples are borderline, and the rest fall clearly outside the extension of the predicate. The whole idea of a 'boundary' does not seem relevant here. Strictly speaking, then perhaps we can say that this sort of predicate does appear to lack sharp boundaries—simply because it has no boundaries at all. But the way in which the predicate appears to lack sharp boundaries seems only loosely connected to the way in which 'bald' appears to lack sharp boundaries.

(3) *Vague terms are susceptible to sorites paradoxes*

We can construct a sorites paradox for the predicate 'bald' as follows. We can take some object which clearly falls within the extension of the predicate—such as a person with 0 hairs. Our first premise is then that a person with 0 hairs is bald. Our second premise is a 'tolerance principle': for any *n*, if a person with *n* hairs is bald, then a person with $n + 1$ hairs is also bald. We can use these two premises to argue that a person with 1 hair is bald, and from there that a person with 2 hairs is bald, and so on. Eventually, we reach the conclusion that a person with 500,000 hairs is bald, and this is clearly false, for a person with 500,000 hairs is a clear case of an object that falls outside the extension of the predicate. This argument is a paradox because it has apparently true premises, seems to be valid, but has an apparently false conclusion.

We cannot construct a sorites paradox in the same sort of way for the predicate '… at … has a resting heart rate of…', for we are missing the item that clearly falls within the extension of the predicate that we need to construct our first premise. Of course sorites paradoxes can be constructed in reverse—so perhaps we could instead start with a premise involving an item that clearly does *not* fall under the extension of the predicate. For example, the first premise could be that I do not currently have a resting heart rate of 20 bpm. The obvious difficulty with this strategy is that to construct the paradox in this way we would need some item which clearly falls under the extension of the predicate in order to reach the (apparently false) conclusion. Another difficulty is in giving a persuasive tolerance principle. We might try: for any animal a, time t, and number n, if a at t has a resting heart rate of n bpm, then a at t has a resting heart rate of $n + 1$ bpm. But this principle doesn't even appear true. Thus it seems that predicates like '… at … has a resting heart rate of …' are not susceptible to sorites paradoxes.

We have surveyed the three features that vague predicates typically display, and found that none of them are displayed straightforwardly by the predicate '… at … has a resting heart rate of…'. If predicates like '… at… has a resting heart rate of…' are vague, then, they are not archetypal vague predicates. However we might think that we could instead focus on some other closely related predicates that *do* display the typical features of vague predicates. For example, take the two-place predicate '… at… has a resting heart rate of at least 100 bpm'. First note that this predicate plausibly has both clear cases and borderline cases: some pairs consisting of an animal and a time fall clearly within its extension; some fall clearly outside its extension; and some are borderline. Second, note that this predicate seems to lack sharp boundaries in much the same way that '… is bald' does: some items are borderline, but the set of borderline items is not sharply bounded. Can we pull off a hat-trick, and show that this predicate also gives rise to sorites paradoxes?

Well, we can produce an item that clearly falls within the extension of the predicate, and so can construct the (apparently true) first premise: e.g. the first premise could be the claim that Scrabble the gerbil currently has a resting heart rate of at least 100 bpm. We can also produce an item that clearly falls outside the extension of the predicate, and so construct the (apparently false) conclusion: e.g. the conclusion could be the claim that Bradley Wiggins the cyclist currently has a resting heart rate of at least 100 bpm. But can we produce a tolerance principle that at least appears to be true? This is tricky. For 'bald', we had a variable—the number of hairs—which we could use to construct the tolerance principle. There we could rely on the natural assumption that whether someone is bald depends on the number of hairs that she has. What could play this role in the sorites paradox for '… at… has a resting heart rate of at least 100 bpm'? What variable does an animal's resting heart rate necessarily depend upon? There doesn't seem to be any obvious answer to this question—except of course this trivial answer: the animal's resting heart rate. But constructing a tolerance principle using this trivial answer is a non-starter, for (no matter how small we make e) it does not even appear to be true for any n that if an animal at a time with a resting heart rate of n has a resting heart rate of at least 100 bpm, then an animal at a time with a resting heart rate of $n–e$ has a resting heart rate of at least 100 bpm. Perhaps we might instead imagine arranging all animals-at-a-time in the relevant order, and trying this principle: if any animal-at-a-time in this series has a resting heart rate of at least 100 bpm, then the animal-at-a-time immediately to its right also has a resting heart rate of at least 100 bpm. But the tricky question here is—what is the relevant order? The idea of course is that the animals-at-a-time should be ranged in order according to their resting heart rate—but how would this be done? The comparative '… has a higher resting heart rate than…' is itself vague.[6] It is not always clear whether one animal-at-a-time has a higher resting heart rate than another, and so there is no clear ordering of all possible animals-at-a-time according to their resting heart rate. In summary it is not at all obvious how we might construct a sorites paradox for this predicate.

Thus whether we stick with our target predicate '… at… has a resting heart rate of…' or consider closely related predicates such as '… at… has a resting heart rate of at least 100 bpm', we find that we are dealing with atypical vague predicates here. In the next section, I consider how we might extend a standard account of vague predicates—supervaluationism—to cover these sorts of predicates. This gives us an account of the (in my view, vague) predicate '… at… has a credence in … of ….'. I see this account as an alternative to the standard imprecise probabilist's account, and in Sect. 5 I examine the ways in which these two accounts differ.[7]

4 Supervaluationism

I start by illustrating the supervaluationist's account using the predicate '… is bald'. The supervaluationist claims that there is a range of admissible ways of making the language precise: in other words, there is a range of admissible precise languages, or 'precisifications'. Perhaps under one such precisification, the boundary for '… is bald' lies at 5739 hairs; another might draw the boundary at 5738 hairs—and there will be many other options. For the supervaluationist, a sentence is super-true if and only if it is true under every precisification; it is super-false if and only if it is false under every precisification; and if it is true under some precisifications, and false under others, then it is neither super-true nor super-false.

Thus, for example, the claim that a person with 0 hairs is bald is true under every admissible precisification, and so this claim is super-true. The claim that a person with

[6] Keefe argues persuasively (Keefe 2000, p. 12), in disagreement with Neil Cooper (Cooper 1995), that comparatives can be vague.

[7] Aidan Lyon (Lyon 2017) also draws on the vagueness literature to give an account of our doxastic states. Lyon draws inspiration from the degree theorists, who claim that sentences have degrees of truth between 0 and 1. Though Lyon's account is inspired by the degree theorists, his account is really a new (and very interesting) extension of the machinery that the degree theorists use to give their account of vagueness. In contrast, I am suggesting that no new account is needed: a theory of vagueness (and in my paper I focus on supervaluationism rather than the degree theory) *already* has the resources to explain the intuitions that seem to drive us away from the orthodox Bayesian account. Many thanks to Dan Lassiter for bringing Lyon's paper to my attention.

500,000 hairs is bald is false under every precisification, so this claim is super-false. The claim that a person with 5739 hairs is bald is true under some precisifications but false under others, so it is neither super-true nor super-false. We can also consider more complex claims, such as the claim that if a person with 5739 hairs is bald, then a person with 5738 hairs is also bald. This is true under every precisification, and so counts as super-true. Similarly, this disjunction comes out as super-true, even though neither disjunct is super-true: either a person with 5739 hairs is bald, or she isn't. More generally, all instances of the law of the excluded middle—and indeed all the laws of classical logic—come out super-true. Furthermore, even though there is no number n such that it is super-true that a person with n hairs is bald but a person of $n + 1$ hairs is not, nevertheless the claim that there is some such n is super-true—for it is true under each precisification. This gives the supervaluationist a solution to the sorites paradox: the tolerance principle is false under every precisification, and therefore super-false.

Can we extend this account to give a supervaluationist account of our predicate '… at … has a resting heart rate (in bpm) of…'? A natural move here is to say that under each precisification of the language, this expression denotes a relation, and the relation consists of ordered triples—each containing an animal, a time, and a number. For each animal and each time, there will be only one number that appears in a triple with that animal and time in the relation: we can then say that the animal and time pair are mapped onto, or assigned, that number. Thus under one admissible precisification of the language, the expression '… at… has a resting heart rate of…' might assign me, right now, the number 75.38: under this precisification, it is true that my resting heart rate is 75.38 bpm. Under a different precisification, the expression will pick out a different relation, which perhaps assigns a different number to me.

Let us say then that under each admissible precisification of the language, the expression '… at … has a resting heart rate of…' refers to a relation that assigns a single number to every animal (with a heart) at every time. Whether a sentence is super-true, super-false, or neither depends on whether the sentence is true under all, none, or some of these precisifications of the language. Thus the sentence 'my resting heart rate now is at least 20 bpm' will come out super-true, because it will be true under every precisification; the sentence 'my resting heart rate now is at least 120 bpm' will come out super-false because it is false under every precisification; and a sentence like 'my resting heart rate now is exactly 75.384' will come out neither super-true nor super-false, because it will be true under some precisifications but false under others.

Here it might be objected that there is an important difference between the supervaluationist account of the vague term '… is bald', and the supervaluationist treatment that I offer of '… at … has a resting heart rate of…'.[8] The supervaluationist claims that there are various admissible precisifications of '… is bald', and (the objector might say) we can give a straightforward intuitive definition for each such precisifation. For example, take the precisification according to which 'bald' means bald_{5739}: we can define this precisification straightforwardly by saying that a person is bald_{5739}

[8]Thanks to the audience at Stirling Philosophy Department for pushing this objection.

iff she has 5739 or fewer hairs on her head. In contrast (the objector might continue), the same does not hold for my supervaluationist treatment of '… at… has a resting heart rate of …'. I claim that there are ways of making this expression precise, but what are these precisifications? How might we define them? The first part of my response is to clarify that it is not just one-place predicates like '… is bald' that can be vague. Predicates with more than one place can be vague too, and given a supervaluationist treatment. To use Keefe's example, each admissible precisification of the predicate '… is a friend of…' has an extension that is a set of ordered pairs (Keefe 2000, p. 159), and similarly, I claim that each admissible precisification of the predicate '… at… has a resting heart rate of …' has an extension that is a set of ordered triples (each triple consisting of a person, a time, and a number). The second part of my response is to clarify that it is not part of the supervaluationist story that there must be a straightforward and intuitive way to characterize each precisification: 'all that is needed for the truth-conditions is the range of precise extensions themselves' (Keefe 2000, p. 159). If the supervaluationist did require a straightforward and intuitive characterization of each precisification, then the supervaluationist could not apply her account to vague terms such as '… is nice'—or indeed '… is a friend of…'. Thus, in response to the objection, I claim that my supervaluationist treatment of the predicate '… at … has a resting heart rate of …' is not importantly different from the supervaluationist treatment of many other expressions.

Having seen how we can give a supervaluationist account of the predicate '… at … has a resting heart rate of…', let us now give a similar supervaluationist account of the predicate '… at… has a credence in … of …'. Under each precisification of the language, this expression will denote some relation. This relation will consist of ordered 4-tuples, each containing a person, a time, a proposition and a number. For any person, time, and proposition, there will be only one number that appears with them in a 4-tuple in the relation, and so we can say that the combination of person, time, and proposition are mapped onto (or assigned) a particular number. Thus for example, under one precisification of the language, the expression will refer to a relation that maps you, now, together with the proposition SARDINES, onto the number 0.342: under this precisification, it is true that your credence in SARDINES right now is 0.352. Under another precisification, it is true that your credence right now in SARDINES is 0.353. Once again, we say that a sentence is super-true iff it is true under every precisification; super-false if false under every precisification, and neither super-true nor super-false if true under some but not all precisifications.[9]

This, then, is one way to defend orthodox Bayesianism: claim that the expression 'credence' is vague and so covered by the supervaluationist's theory of vagueness.[10]

[9]It follows on this account that for any pair of propositions P and Q, and any rational agent, the following disjunction will be super-true (because true under all precisifications): either the agent has a higher credence in P than in Q, or the agent has a higher credence in Q than in P, or the agent has the same credence in P as in Q. As Richard Dietz pointed out to me, this is in conflict with the views of some theorists (Keynes 1921)—though we can soften the conflict by adding that on the supervaluationist account it may be that none of the disjuncts are super-true.

[10]I claim that the expression 'credence' is vague, but I do not claim that it is vague along every dimension. For example, take the predicate '… is a credence function'. At least on some views, this

The orthodox Bayesian's claim—that every rational agent has a precise credence in every proposition entertained—then emerges as super-true (because it is true under every admissible precisification), but it does not have the unwelcome consequences that we might expect. For example, on this view, orthodox Bayesianism is compatible with there being no particular n such that it is super-true that your credence in SARDINES is n. This, then, is one way that the orthodox Bayesian could reconcile her account with our intuitions. In the next section, I turn to an alternative: the imprecise probabilism account. This influential account is also sometimes referred to as a type of supervaluationism (Hajek 2003, pp. 277–8; van Fraassen, Vague Expectation Value Loss 2006, p. 483), but it is an account inspired by supervaluationism, rather than a mere application of the standard theory. I describe this imprecise probabilism account in the next section, and then compare it with the supervaluationist account that I have given above.

5 Imprecise Probabilism

On the orthodox Bayesian view, a rational agent's credal state can be represented by a single credence function. Many theorists, including (Jeffrey 1983; van Fraassen 1990; Joyce 2010), deny this and claim instead that a rational agent's credal state can be represented by a *set* of precise credence functions. The credence functions in the set will be such that whatever holds of the agent's credal state holds for every credence function in this set: or as Joyce puts it, '[f]acts about the person's opinions correspond to properties common to all the credence functions in her credal set' (Joyce 2010, p. 287). Thus, for example, suppose that your credence in SARDINES is greater than 0.2: in that case every credence function in the set representing your credal state must assign a value of at least 0.2 to SARDINES. Suppose also that your credence in SARDINES is less than 0.8: in that case, every credence function in the set representing your credal state must assign a value of less than 0.8 to SARDINES. It is natural to think of your credence in SARDINES as a range—given by the range of numbers that the credence functions in the set representing your credal state assign to SARDINES. We can also consider more complex claims about your credal state. For example, it may be that your credence in SARDINES is greater than your credence in EXTREME SARDINES, where EXTREME SARDINES is the claim that my

predicate is precise, for everything either is a credence function (i.e. a function—perhaps conforming to certain axioms, depending on your preferred definition of 'credence function—from propositions to numbers between 0 and 1) or it isn't. My claim is rather that it can be vague whether a particular agent at a time has a particular credence function. Furthermore, the idea is not simply that it might be vague whether a particular agent has *any* credence function (perhaps because it is vague whether the agent is capable of being in a belief state), or that it might be vague what relation the agent stands into the function (i.e. whether it is an epistemic relation rather than some other closely related relation). The idea is that it can be vague *which* credence function is the credence function of a particular agent at a specific time. Thanks to the audience at the philosophy department in Sterling for pushing me to clarify these points.

neighbour has more than ten tins of sardines in her cupboard. This claim about your credal state will obtain iff every precise credence function in the set representing your credal state assigns a higher number to SARDINES than to EXTREME SARDINES.

Having given this brief overview of imprecise probabilism, I turn to a comparison with the supervaluationist account given in the previous section. The fundamental difference between the accounts it that for the supervaluationist, the orthodox Bayesian theory is true but expressed in vague language; whereas for the imprecise probabilist, the orthodox Bayesian theory is false and should be replaced with imprecise probabilism, on which we have alternative models that represent epistemic states precisely. In the following three sections I compare the accounts by looking at three different features: I consider how each account handles complex claims; then I consider how a decision theory can be constructed on each account; and finally I look at how each account can handle higher-order vagueness (or its analogue).

6 Complex Claims

Here is an important difference between the two accounts. The imprecise probabilist works with sets of precise credence functions, and each such set of credence functions is designed simply to represent the credal state of a particular agent at a particular time.[11] In contrast, the supervaluationist works with a set of precise languages, and these languages can be used to say all sorts of things about all sorts of topics.

For the supervaluationist, each precisification of our language contains a precisification of the expression ‘credence’. We can interpret a precisification of ‘credence’ (as we did earlier) as a 4-place predicate that we can see as assigning numbers to various combinations of individual, time, and proposition. For example, a given precisification of ‘credence’ might assign the number 0.342 to the combination of me, the present time, and the proposition SARDINES, and also assign the number 0.341 to the combination of you, the present time, and the proposition SARDINES. To put it another way, under this precisification of the language, it is true that my credence right now in SARDINES is 0.342, and also true that your credence right now in SARDINES is 0.341. Thus these claims about both your and my credal states each get a truth-value under this single precisification of the language. Similarly, claims about the relation between your credal state and mine get a truth-value under each precisification of the language. For example, take the claim that my credence (now) in SARDINES is greater than yours (now). This claim will have a truth-value under each precise language: for example, on the precisification just mentioned, under which it is true that my credence right now in SARDINES is 0.342, and true that your credence right now in SARDINES is 0.341, it will be true that my credence in

[11]On some versions of imprecise probabilism, we represent an agent’s mental state with a set of *pairs*, each pair consisting of a credence function and a utility function. A set of such pairs represents more than just the agent’s credence function—but there is still an enormous contrast with the supervaluationist’s set of precise languages, which can represent far more than the mental state of a single agent at a time.

SARDINES is greater than yours. In this way claims relating multiple credal states will have a truth value under any given precisification. And on the supervaluationist account, we can go on to say that a claim relating multiple credence functions will be super-true if and only if it is true under every precisification, super-false if and only if it is false under every precisification, and otherwise neither super-true nor super-false. Thus, on the supevaluationist account, there is no difficulty in explaining how truth-values are assigned to claims relating different credal states.

What about the imprecise probabilist's account? Well, here we do not have a set of precise languages, but rather sets of precise credence functions. Your current credence function is represented by one such set, and my current credence function is represented by some other such set. It makes no sense to talk about what my credence is under one of the precise credence functions that represent your credal state, or vice versa. The truth-value of claims about your credal depend on the set of credence functions that represent your credal state; the truth-value of claims about my credal state depend on the set of credence functions that represent my credal state. This leaves the imprecise probabilist with a question: what settles whether a claim holds when that claim is about the relationship between our credal states?

The imprecise probabilists have not, as far as I know, generally agreed on an answer to this question, but there are at least two sorts of answers that the imprecise probabilist could give:

(i) The imprecise probabilist could say that whether my credence in SARDINES is greater than yours depends simply on our credal ranges in these two propositions. For example, the condition could be that the highest value in my credal range must be greater than the highest value in your credal range. Alternatively—or additionally—it might be required that the lowest value in my credal range must be greater than the lowest value in your credal range. An alternative and more demanding requirement could be that the lowest value in my credal range must be greater than the highest value in your credal range.[12]

(ii) The imprecise probabilist could say that whether my credence in SARDINES is greater than your credence in SARDINES depends on whether a particular sort of mapping relation holds between the credence functions in each of our sets. To understand this idea, imagine trying to one-to-one map each credence function in your set to some corresponding credence function in my set—the only criterion being that for each credence function in your set, the number assigned to SARDINES must be less than the number assigned to SARDINES by the corresponding credence function in my set. If some such mapping

[12]This leaves open the question of when we should say that my credence in SARDINES is equal to yours—and this is a further issue for the imprecise probabilist to resolve. One option here is to say that all that is required is that my credence in SARDINES must be neither lower nor higher than yours—which (on the 'demanding requirement' in the text above) would mean just that there needs to be some overlap between your credal range and mine. Perhaps, though, this makes equality implausibly easy to come by. In contrast, we might rule that my credence in SARDINES is equal to yours only if our credal ranges are identical, and this may make equality implausibly hard to come by. There is an interesting parallel here with Joshua Gert's account of value (Gert 2004), and I am grateful to Richard Dietz for pointing me to this account.

> is possible, then my credence in SARDINES is higher than yours. We can similarly define what it is for my credence in SARDINES to be lower than yours, and it is natural to extend the account to define what it is for my credence in SARDINES to be equal to yours—which would be for there to be a function that maps each credence function in your set to some corresponding credence function in my set that assigns the very same number to SARDINES.[13,14]

The question that we have put to the imprecise probabilist may not seem particularly pressing. Perhaps the whole idea of my credence being greater than yours in SARDINES is rather obscure, given that our credences in this claim are intuitively imprecise. What would it mean for my credence function to be higher than yours in this context? But there is a related question that does seem more pressing, and this concerns comparisons of the same person's credal state across time. For example, suppose that you come to learn NOT EXTREME SARDINES—which is the claim that my neighbour does *not* have more than ten tins of sardines in her cupboard. Intuitively, on learning this your credence in SARDINES ought to decrease, for one of the ways that SARDINES could be true has just been eliminated. But what makes it the case that your credence in SARDINES is higher before gaining this evidence than it is after? It is not immediately obvious how the imprecise probabilist should answer this question, but we may be able to construct an answer by thinking about the imprecise probabilist's approach to rational updating, which I turn to now.

On the orthodox Bayesian view, a rational agent updates by conditionalizing on any new evidence that she encounters. Thus if a rational agent encounters (only) some new piece of evidence E, then her new credence function will be identical to her old credence function conditionalized on E. What does the imprecise probabilist require of a rational agent who encounters only E? For the imprecise probabilist, the credal states of the agent at the earlier and later times will each be represented by a set of credence functions (rather than a single credence function). What should the relationship between these two sets be, if the agent has updated on new evidence E, as rationality demands? The answer the imprecise probabilists gives[15] is that there should be a particular sort of one-to-one mapping relation between the two sets: we should be able to map each function from the earlier set onto some function from the later set, such that the function from the later set is identical to the corresponding function from the earlier set conditionalized on E.[16]

---

[13] It follows on this account that for my credence in SARDINES to be greater than yours (or lower than yours or equal to yours) the sets of credence functions that represent each of our credal states need to agree in cardinality. This seems like a very strong constraint, and this may be an objection for an imprecise probabilist who takes this option. Thanks to Richard Dietz for pointing this out.

[14] I suspect that—given certain natural constraints—this second option (ii) will collapse into some version of the first option (i). We will see the point of introducing the mapping relation below when discussing conditionalization.

[15] This is at least the answer usually given—but other updating rules are possible. See (Joyce 2010, pp. 292–3) for a discussion of an alternative.

[16] 'If a person in credal state C learns that some event D obtains (and nothing else), then her post-learning state will be $\mathrm{C}_D = \{c(\bullet|D) = c(X) \cdot [c(D|X)\,/c(D)] : c\in\mathrm{C}\}$' (Joyce 2010, p. 287).

In describing updating on the imprecise probabilist's account, it is tempting to think of the individual credence functions enduring between the earlier and the later time—as though when the agent encounters E, what happens is that each individual precise credence function is updated. This impression is reinforced by the common use of a (vivid and helpful) metaphor, in which we think of the agent's credal state as a set of 'avatars' (Bradley 2009) or 'committee members' (Joyce 2010), each with a precise credence function that gets updated as the agent whose credal state they represent gains evidence. Here it is natural to think of each avatar enduring a change in its credence function—but this was always intended just as a metaphor. If we focus instead on the sets of credence functions themselves, then it is clear that it does not make sense to think of a single credence function changing its assignment of values over time: functions that assign different values are simply different functions. Thus, in a case of rational updating, there is no identity relation that connects each credence from the earlier set with some credence in the later set. On the imprecise probabilism view, then, the relationship that must hold for an agent to have updated rationally in response to new evidence must be as I described above: the condition is that there must be some one-to-one mapping relation between the two sets such that each function from the earlier set is mapped onto a function from the later set, which assigns every proposition whatever number the function from the earlier set assigned to that proposition conditional on E.

Given that this mapping relation is what is needed for an agent to count as having updated rationally, it is natural to also use mapping relations to explain relations between different credal states more generally. Thus we can say that for your credence in SARDINES to decrease after learning EXTREME SARDINES is for a certain mapping relation to hold between the credence functions in the set that represent your earlier credence function and those in the set that represent your later credence function: there must be some one-to-one mapping of each credence function from the earlier set onto some credence function in the later set, such that the credence function from the earlier set assigns a higher value to EXTREME SARDINES than the credence function in the later set. And we can tell a similar story about what must be the case for your credence in SARDINES to be lower than mine—along the lines of (ii) above. Thus it is possible for the imprecise probabilist to give an account of these sorts of complex statements, but the account does not follow automatically from the core claims of imprecise probabilism.[17]

For the supervaluationist, the story is pretty straightforward. Under any given precisification, it will be either true or false that your credence function in SARDINES is lower than mine; that my credence in SARDINES has decreased since learning EXTREME SARDINES; or that on learning some evidence I have conditionalized as rationality demands. Whether these claims are super-true, super-false, or neither simply depends on whether the claims are true under all, none, or some of the precise languages. Thus whether we go for a supervaluationist account, or imprecise probabilism, we can make sense of claims that relate one credal state to another—

[17] See footnotes 13 and 14 for some of the difficulties in giving such accounts.

but while this follows automatically for the supervaluationist, the details need to be bolted on to imprecise probabilism.

7 Decision Theory

As we have seen, the supervaluationist works with a set of precise languages, which can be used to say things about all sorts of topics; in contrast, the imprecise probabilist works with sets of credence functions, and each set is used just to represent a single agent's credal state at a time. We have seen how this leads to a difference in the way that the two accounts handle claims about relations between the credences of different people, or about the relations between the same person's credence at two different times. It also leads to a difference in the way that decision theory works on the two accounts.

The imprecise probabilists have put forward many different decision theories, but here I just briefly describe two sample theories[18]:

1. *Caprice* (Weatherson 2008): on this theory, an agent is rationally required to maximise expected utility relative to at least *one* credence function in the set that represents her credal state. Thus, for example, suppose that I offer you a bet on SARDINES, whereby you pay out £0.35 and get £1 back iff SARDINES. Let us assume that you value only money, and value that linearly. Then you are permitted to accept this bet iff there is a credence function within the set that represents your credal set that assigns SARDINES a value of at least 0.35. Similarly, you are permitted to reject this bet iff there is a credence function within the set that assigns SARDINES a value of no more than 0.35. There may be many bets that you are rationally permitted to either accept or reject (at your caprice).
2. *Maximin* (Gärdenfors and Sahlin 1982) (Gilboa and Schmeidler 1989): On this theory, an agent is rationally required to maximise minimum expected utility. To understand what this means, let us again suppose that I offer you a bet on SARDINES, whereby you pay out £0.35 and get £1 back iff SARDINES is true. You have the option of either accepting the bet or rejecting it. We can start by calculating the minimum expected utility of accepting the bet. To do this, we calculate the expected utility of accepting the bet relative to each credence function in your credal set: the lowest expected utility that we get from this process is the minimum expected utility of accepting the bet. We then calculate the minimum expected utility of rejecting the bet in a similar way. Thus the actions available to you (accepting the bet, rejecting the bet) each have a corresponding minimum expected utility, and if one has a higher minimum expected utility than the other, then you are required to choose that action. More generally, in any choice situation you are rationally required to select from amongst those actions with the maximum minimum expected utility.

[18] I focus on normative theories, but there are also attempts to produce descriptive theories.

Various problems face both of these decision rules, and I discuss some of these problems elsewhere (Mahtani 2016). But the point that I want to note here is that both of these decision rules refer to the *set* of credence functions that represent the agent's credal state.

For the supervaluationist, in contrast, the most natural thought is that a given decision rule will be stated in the object language, and will make no reference to the set of admissible precsificiations. For example, the supervaluationist might endorse the simple decision rule MEU, according to which a rational agent always chooses an action that maximises expected utility. On the supervaluationist account, different precisifications of the language precisify 'credence' differently, and so the actions that can truly be said to maximise expected utility, and therefore be rational, can vary across precisifications. For example, suppose that I offer you a bet whereby you pay out £0.35 and get £1 back iff SARDINES obtains. We assume as before that you value only money, and value it linearly. Should you then accept or reject this bet? Which action has higher expected utility? Well, the expected utility of the actions depend on your credence in SARDINES, and 'credence' is vague and so can be made precise in different ways. Under some precisification, perhaps it is true that your credence in SARDINES is 0.351, and so the expected utility of accepting the bet is greater than the expected utility of rejecting it, and so it would be rational to accept the bet. Under another precisification, perhaps it is true that your credence in SARDINES is 0.349, and so the expected utility of accepting the bet is less than the expected utility of rejecting it, and so it would not be rational for you to accept the bet. Thus it may be true under one precisification that an action is rational, but false under another. It is super-true that an action is rational only if the claim that it is rational is true under every precisification; it is super-true that an action is not rational only if the claim that it is not rational is true under every precisification. If the claim that an action is rational is true under some but not all precisifications of the language, then it is neither super-true that the action is rational nor super-true that the action is not rational: we can then say that it is indeterminate whether the action is rational. Let us call this the supervaluationist's simple decision theory.[19]

Here supervaluationism seems to give us a simpler decision theory than imprecise probabilism.[20] The imprecise probabilist needs to construct some new sort of decision theory to accommodate the fact that an agent's credal state is represented by a set of credence functions rather than a single credence function. In contrast, the supervaluationist can stick with decision rules developed by the orthodox Bayesian, such as MEU. But does the supervaluationist's simple decision theory have the right implications?

One problem for the supervaluationist's simple theory is that in some cases an agent may be faced with a decision problem in which there is no action such that it is super-true that that action is rational. To see this, consider again the decision problem above in which you must choose between accepting and rejecting the bet for which

[19]Robert Williams briefly explores a related idea in (Williams 2014, pp. 25–26).

[20]Of course, much depends on whether this simpler decision theory is plausible—a question that I explore below.

you pay out £0.35 and get £1 back iff SARDINES is true. Is it super-true that it is rational to accept the bet? Well, if under some admissible precisification it is true that your credence in SARDINES is 0.349, then under this precisification it is false that it would be rational to accept the bet, and so it is not super-true that it would be rational to accept the bet. Is it then super-true that it is rational to reject the bet? Well, if under some admissible precisification it is true that your credence in SARDINES is 0.351, then under this precisification it is false that it would be rational to reject the bet, and so it is not super-true that it would be rational to reject the bet. Thus it may be that neither the claim that it is rational to accept the bet, nor the claim that it is rational to reject the bet, come out as super-true. Whichever action you perform, then, it will not be super-true that your action is rational. This looks like a problem: it seems that in any given decision problem there should be at least one action such that it is super-true that choosing this action is rational.

Another problem for the supervaluationist's simple decision theory is the phenomenon of 'ambiguity aversion'. This can be demonstrated with the 'Ellsberg Paradox' (Ellsberg 1961). Suppose that you know that an urn contains 30 red balls, and 60 balls that are either black or yellow: you do not know the ratio of black to yellow balls. You are about to draw out a ball at random, and are given a choice between these two options: (a) you receive £100 if you draw a red ball; (b) you receive £100 if you draw a black ball. Now suppose instead that you are given the choice between these two options: (c) you receive £100 if you draw a red or yellow ball; and (d) you receive £100 if you draw a black or yellow ball. Many apparently rational people prefer (a) to (b), but prefer (d) to (c). This is the phenomenon of 'ambiguity aversion': as the imprecise probabilist would put it, a rational agent may prefer an action where every credence function in the set that represents her credal state assigns this action the same expected utility, over an action where the expected utilities assigned differ—all else being equal. Some of the decision rules put forward by the imprecise probabilist are designed to explain this phenomenon: the rule maximin outlined above is one such rule. But can the phenomenon be accommodated on the supervaluationist's account?

On the simple supervaluationist's decision theory that we are exploring here, under each precisification of the language the agent has some particular credence in (RED) the ball drawn being red, some particular credence in (BLACK) its being black, and some particular credence in (YELLOW) its being yellow. Presumably on some precisifications of the language, the agent has a higher credence in RED than BLACK, while on others the agent has a higher credence in BLACK than RED. Thus under some precisifications of the language (those on which the agent has a higher credence in RED than BLACK) the agent prefers (a) to (b) and (c) to (d), while under other precisifications of the language (those on which the agent has a higher credence in BLACK than RED), the agent prefers (b) to (a) and (d) to (c). But this leaves some claims about the agent's preferences—e.g. the claim that the agent prefers (a) to (b)—indeterminate, i.e. neither super-true nor super-false. And it leaves the complex claim—that the agent prefers (a) to (b) and (d) to (c)—super-false. And this just doesn't seem right: many rational people report an apparently definite preference for (a) over (b) and (d) over (c).

To handle these two problems, the supervaluationist can move away from her simple decision theory to a theory that involves the terms ‘super-true’/‘super-false’, and ‘admissible precisifications’. These terms are part of the supervaluationist’s meta-language—that is, the language that the supervaluationist uses to express her theory. Thus, if the decision theory uses these terms, then the decision theory will also be in the meta-language. This may seem strange, because surely the meta-language is designed to express supervaluationism rather than to express a decision theory. However this is not really such a big departure for supervaluationism, for supervaluationists typically ascend to the meta-language to justify their position, and this can involve showing how supervaluationism can successfully predict and explain verbal behaviour. For example, the supervaluationist can explain why, when confronted with a borderline red object and asked whether it is red, people will often refuse to assert either that the object is red or that object is not red, perhaps saying instead that there is no fact of the matter: the supervaluationist can explain this by claiming that this is how people do (and perhaps should) respond in cases where neither a statement nor its negation are super-true. Thus the supervaluationist typically ascends to the meta-language to explain and predict verbal behaviour, and so it is no great jump for the supervaluationist to ascend to the meta-language in order to explain and predict behaviour in general—i.e. in order to give a decision theory.[21]

By ascending to the meta-language, we can express some alternative decision theories that the supervaluationist could adopt. For example, the supervaluationist could claim that an action is rationally permissible if and only if that there is at least one precisification under which it is true that the action maximises expected utility. This gives us a decision rule parallel to the imprecise probabilist’s rule ‘caprice’. Alternatively, the supervaluationist could claim that an action is rationally permissible if and only it has the highest minimum expected utility—where an action’s minimum expected utility is the lowest expected utility that it is assigned under any precisification. This gives us a decision rule parallel to the imprecise probabilist’s rule ‘maximin’, which can explain and predict ambiguity aversion.

In summary, then, the supervaluationist can offer a range of different decision theories. There is the supervaluationist’s simple decision theory, which does not have a parallel in the imprecise probabilist’s account. In addition, by ascending to the metalanguage the supervaluationist can offer a range of decision rules, which parallel those offered by the imprecise probabilist. Thus the supervaluationist has resources at least as good as imprecise probabilism to offer a plausible decision theory.

[21] An alternative for the supervaluationist may be to use the term ‘determinately’ (or ‘definitely’) to state her decision theory. To define this term: for any claim *p*, if the claim *p* is super-true, then the claim ‘determinately *p*’ is true under every precisification; if the claim *p* is not super-true, then the claim ‘determinately *p*’ is false under every precisification. The idea is that these terms are part of the object language, and allow us to express what we would otherwise have to express in the meta-language. It may be (though I have not shown this) that whatever decision rule can be expressed in the meta-language can equally well be expressed in the object language, supplemented with ‘determinately’.

8 Higher-Order Vagueness

Supervaluationism faces the problem of higher-order vagueness. Recall that intuitively there is no sharp boundary between the possible objects that are bald and those that are not bald: this is what motivates supervaluationists to construct their account. But the supervaluationist's account seems to avoid positing one sharp boundary, only to posit two more: one marking off the objects that it is super-true to describe as bald from the borderline cases, and another marking off the objects that it is super-false to describe as bald from the borderline cases. Intuitively there are no sharp boundaries here either—and in fact 'bald' does not draw any sharp boundaries whatsoever. How can the supervaluationist handle this?

A convincing response to this problem is to claim that the metalanguage—the language in which the theory of supervaluationism is expressed—is itself vague. In particular, the expression 'admissible precisification' is vague: there is no sharp boundary between the precisifications that are admissible and those that are not, and so it follows that there is no sharp boundary between those claims that are super-true/-false and those that are not. The supervaluationist can then give the same account of the vagueness in the metalanguage as she gave of the vagueness in the object language: there are many admissible precisifications of the metalanguage, and whether a statement in the metalanguage is super-true will depend on whether the sentence is true under each precisification of the metalanguage.[22] Of course we will then want to say that the meta-meta-language is also vague, and so on, and we will end up with an infinite hierarchy of vague metalanguages—but this is not obviously problematic (Keefe 2000, pp. 202–213). Thus the supervaluationist has at least one promising response to the problem of higher-order vagueness.

The imprecise probabilist faces a problem analogous to higher-order vagueness (Maher 2006; Kaplan 2010; Rinard forthcoming). For the imprecise probabilist, the problem can be put as follows. Intuitively you do not have a precise credence in SARDINES, for you do not know what your credence is in this claim, and there doesn't seem to be anything that makes it the case that your credence is one value rather than another. The imprecise probabilist handles this problem by claiming that your credal state is represented by a set of credence functions, and these will assign between them a range of values to SARDINES. But what is this range, exactly? What is the lowest number in this range, and what is the highest? The problem with the orthodox Bayesian theory seems to resurface in a different form. For you don't know what numbers form the upper and lower bound of your credal range. And there doesn't seem to be anything that could make it the case that these upper and lower bounds are any particular numbers. How can the imprecise probabilist handle this problem?

One option for the imprecise probabilist is to claim that the account of imprecise probabilism is given in a vague language, and then to use something like supervalu-

[22] This sentence—which gives an account of the vagueness of the meta-language—is itself in the meta-meta-language, and the expression 'super-true' cannot be assumed to have the very same meaning in the meta- and meta-meta-languages.

ationism to give an account of this vagueness. Thus the imprecise probabilist could claim perhaps that it is vague whether a particular credence function belongs in the set that represents an agent's credal state. This approach is not very appealing, however, because it makes the imprecise probabilist's position quite complicated, with different sorts of accounts needed to handle the original problem and the 'higher-order' problem. The imprecise probabilist would have to claim both that the standard Bayesian model should be replaced by an alternative model, and then that the account describing the alternative model is vague and should be given a supervaluationist treatment. It would be simpler to give a thorough-going supervaluationist account, as I have done, and drop the claim that the standard Bayesian model needs to be replaced.

As an alternative, the imprecise probabilist might try to iterate her original account at higher-levels, just as the supervaluationist gives a supervaluationist treatment of vagueness at every level. To work out how this might look, recall that the imprecise probabilist recommends that we reject the standard Bayesian account on which an agent's epistemic state is modelled with a single credence function, in favour of an account on which an agent's epistemic state is modelled with a set of credence functions. Then, in response to a problem analogous to higher-order vagueness, the imprecise probabilist might claim that this account, which models an agent's credal state with a set of credence functions, should in turn be replaced by an account that models an agent's credal state with a set of sets of credence functions. This account will then be replaced by a still more complex account—and so on. But this response lacks the appeal of the supervaluationist's response to higher-order vagueness. The imprecise probabilist originally *rejected* the standard Bayesian account in favour of her alternative. Thus, presumably each iteration of the imprecise probabilist's account should similarly be rejected in favour of the next more complex account—meaning that every account in this series should be rejected.[23] Perhaps we should accept an account on which an agent's credal state is modelled by an infinitely long series of sets of credence functions—but what use could we have for such an account? The supervaluationist's response to higher-order vagueness does not suffer from the same problem. The original supervaluationist account of vagueness in the object language is perfectly in order as it is, and the phenomenon of higher-order vagueness gives us no reason to reject it. The account is expressed in vague language, but that is no reason to reject it: the supervaluationist can maintain that it is the correct account of vagueness in the object language. It is only if we want an account of vagueness in the meta-language that we need to ascend into a meta-meta-language to give this account.

The supervaluationist then has a plausible response to the problem of higher-order vagueness. The imprecise probabilist faces an analogue of this problem, and I have looked at some ways in which the imprecise probabilist might respond—but there are problems with each option considered.

[23]This criticism is related to the criticism that Keefe levels against a version of the degree theory of truth (Keefe 2000, pp. 117–120).

9 Conclusion

Orthodox Bayesiansim—which rules that every rational agent has a precise credence in every proposition that she entertains—is counterintuitive. Imprecise probabilism has proved a popular refinement of orthodox Bayesian, but I have argued that there is a viable alternative. The alternative involves recognizing that the expression 'credence' is vague—just as 'resting heart rate' is vague. We can then use supervaluationism, more or less in its standard form, to give an account of these expressions. Using this account, I have offered a simple way of handling sentences relating the credences of different people, or of the same person at different times; I have also suggested some options for constructing a decision theory; and I have shown how we can make use of a convincing supervaluationist approach to higher-order vagueness.

Part of the appeal of replacing imprecise probabilism with a supervaluationist account is that the supervaluationist account comes for free. Vagueness is a pervasive phenomenon, and we need to give an account of it. Supervaluationism is arguably the best account of vagueness that we have, and so every vague term should be given a supervaluationist treatment. That the expression 'credence' is vague is hardly controversial: pretty much all the terms in our language are vague, and there is no particular reason why 'credence' should be an exception. All of this implies that—for reasons unconnected to epistemology—we should give a supervaluationist treatment of the expression 'credence'. I have argued that once we give a supervaluationist treatment of this term, we get everything that imprecise probabilism offers and more, and so imprecise probabilism should be abandoned as unnecessary.

Acknowledgements Many thanks to audiences at the 5th LSE conference on Philosophy of Probability (2016), the Experience and Updating workshop in Bochum (2016), the Visiting Speakers Seminar Series in Stirling (2017), and the Knowledge, Belief, Evidence Conference at Oxford (2018), as well as to all the people who gave me invaluable feedback on this work. I'm also very grateful to the Leverhulme Trust, who supported this research.

References

Bradley, R. (Forthcoming). *Decision theory with a human face.*

Bradley, R. (Forthcoming). Ellsberg's paradox and the value of chances. *Economics and Philosophy*.

Bradley, R. (2009). Revising incomplete attitudes. *Synthese, 171*(2), 235–256.

Buckler, J. M. (1978, Sep). Variations in height throughout the day. *Archives of Diseases in Childhood, 53* (9), 762.

Cooper, N. (1995). Paradox lost: Understanding vague predicates. *International Journal of Philosophical Studies, 3,* 244–269.

De Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg & H. E. Smokler (Eds.), *Studies in Subjective Probability* (pp. 93–158). New York: Wiley.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics, 75*(4), 643–669.

Gärdenfors, P., & Sahlin, N.-E. (1982). Unreliable probabilities, risk taking and decision making. *Synthese, 53,* 361–386.

Gert, J. (2004). Value and parity. *Ethics, 114*(3), 492–510.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics, 18,* 141–153.

Hajek, A. (2003). What conditional probability could not be. *Synthese, 137*(3), 273–323.

Horgan, T. (Forthcoming). Troubles for formal epistemology. *Res Philosophica*.

Jeffrey, R. (1983). Bayesianism with a Human Face. In J. Earman (Ed.), *Minnesota studies in the philosophy of science, Volume 10: Testing scientific theories* (pp. 133–156). Minneapolis: University of Minnesota Press.

Joyce, J. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives, 24,* 281–323.

Kaplan, M. (2010). Defense of modest probabilism. *Synthese, 176*(1), 41–55.

Keefe, R. (2000). *Theories of vagueness*. Cambridge: Cambridge University Press.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan & Co., Ltd.

Lyon, A. (2017). Vague credence. *Synthese, 194*(10), 3931–3954.

Maher, P. (2006). Review of David Christensen. Putting logic in its place: formal constraints on rational belief. *Notre Dame Journal of Formal Logic, 47,* 133–149.

Mahtani, A. (2016). Imprecise probabilities and unstable betting behaviour. *Noûs*.

Palatini, P. (2009). Recommendations on how to test resting heart rate. *Medicographia, 31*(4), 414–420.

Rinard, S. (Forthcoming). Imprecise probability and higher order vagueness. *Res Philosophica*.

Unger, P. (1979). There are no ordinary things. *Synthese, 41*(2), 117–154.

van Fraassen, B. C. (1990). Figures in a probability landscape. In M. Dunn & A. Gupta (Eds.), *Truth or Consequences* (pp. 345–356). Dordrecht: Kluwer.

van Fraassen, B. C. (2006). Vague expectation value loss. *Philosophical Studies, 127,* 483–491.

Weatherson, B. (2008). Decision making with imprecise probabilities. *Manuscript*.

Williams, R. (2014). Decision-making under indeterminacy. *Philosophers' Imprint, 14* (4).

Problems of Precision in Fuzzy Theories of Vagueness and Bayesian Epistemology

Nicholas J. J. Smith

Abstract A common objection to theories of vagueness based on fuzzy logics centres on the idea that assigning a single numerical degree of truth—a real number between 0 and 1—to each vague statement is excessively precise. A common objection to Bayesian epistemology centres on the idea that assigning a single numerical degree of belief—a real number between 0 and 1—to each proposition is excessively precise. In this paper I explore possible parallels between these objections. In particular I argue that the only good objection along these lines to fuzzy theories of vagueness does not translate into a good objection to Bayesian epistemology. An important part of my argument consists in drawing a distinction between two different notions of degree of belief, which I call dispositional degree of belief and epistemic degree of belief.

1 Introduction

A common objection to theories of vagueness based on fuzzy logics centres on the idea that assigning a single numerical degree of truth—a real number between 0 and 1—to each proposition is excessively precise. A common objection to Bayesian epistemology centres on the idea that assigning a single numerical degree of belief—a real number between 0 and 1—to each proposition is excessively precise. There is a striking surface similarity between these objections. I have discussed the objection to fuzzy theories of vagueness elsewhere (Smith 2008, 2011). I have argued that it is a powerful objection to a basic fuzzy theory of vagueness—but I have proposed a more sophisticated fuzzy theory, which avoids the objection. This paper discusses the argument against Bayesian epistemology. In particular it asks whether there is a good argument against Bayesian epistemology *of the same sort* as the good argument against the basic fuzzy theory of vagueness. I'll argue that there isn't: despite the surface similarity, a line of argument that is powerful in one context fails in the other. The main purpose of this investigation is to clarify the conceptual terrain—not to

N. J. J. Smith (✉)
Department of Philosophy, The University of Sydney, Sydney, NSW, Australia
e-mail: nicholas.smith@sydney.edu.au

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*,
Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_3

argue for (or against) Bayesianism. There is (as we'll see) another kind of argument against Bayesian views that works in a different way from the argument against fuzzy theories of vagueness—and although I'll *distinguish* this other kind of argument, I shall not *assess* it here.

The paper proceeds as follows. Section 2 introduces the basic fuzzy theory of vagueness and the objection that it involves artificial precision. Section 3 introduces Bayesian epistemology and the objection that *it* involves false precision. Section 4 introduces an important distinction between two kinds of degree of belief, which I call *dispositional* degree of belief and *epistemic* degree of belief. Section 5 explores whether there is a good argument, to the conclusion that Bayesian epistemology involves false precision, that runs along the same lines as the argument that fuzzy theories of vagueness involve artificial precision: I'll argue that there isn't. Section 6 turns to actual arguments in the literature (to the conclusion that Bayesian epistemology involves false precision). In light of the discussion in Sect. 5, we'll see that some of these arguments fail—while others are of a quite different nature from the artificial precision objection to fuzzy theories of vagueness.

2 Fuzzy Theories of Vagueness

A classical or crisp set is a collection of objects, where each object is either (definitely) in the set or (definitely) not in it. A fuzzy set is a collection of objects, where objects may be in the set to greater or lesser degrees. These degrees are modelled by real numbers from 0 (representing definite or total non-membership) to 1 (representing definite or total membership). A fuzzy set—or more precisely, a fuzzy subset S of some background universe U—is then a function from U to the real unit interval $[0, 1]$. For any object a in U, the number to which a is mapped by the function is a's degree of membership in S.

The core idea of fuzzy theories of vagueness is that the extensions of vague predicates such as 'tall' and 'heavy' are fuzzy sets and (hence) that sentences may have intermediate degrees of truth. For example, the degree of truth of the sentence 'Bob is tall' will be the same as Bob's degree of membership in the fuzzy set of tall things. This degree is a real number in the interval [0, 1]. So in the fuzzy framework, reals in [0, 1] do double duty as degrees of membership (of objects in sets) and degrees of truth (of sentences).

A classic objection to fuzzy theories of vagueness is that they involve excessive or artificial precision:

> [Fuzzy logic] imposes artificial precision…[T]hough one is not obliged to require that a predicate either definitely applies or definitely does not apply, one *is* obliged to require that a predicate definitely applies to such-and-such, rather than to such-and-such other, degree (e.g. that a man 5 ft 10 in tall belongs to *tall* to degree 0.6 rather than 0.5) (Haack 1979, 443).

> One immediate objection which presents itself to [the fuzzy] line of approach is the extremely artificial nature of the attaching of precise numerical values to sentences like '73 is a large number' or 'Picasso's *Guernica* is beautiful'. In fact, it seems plausible to say that the nature of vague predicates precludes attaching precise numerical values just as much as it precludes attaching precise classical truth values (Urquhart 1986, 108).

> [T]he degree theorist's assignments impose precision in a form that is just as unacceptable as a classical true/false assignment. In so far as a degree theory avoids determinacy over whether a is F, the objection here is that it does so by enforcing determinacy over the *degree* to which a is F. All predications of 'is red' will receive a unique, exact value, but it seems inappropriate to associate our vague predicate 'red' with any particular exact function from objects to degrees of truth. For a start, what could determine which is the correct function, settling that my coat is red to degree 0.322 rather than 0.321? (Keefe 1998, 571).

The objection to the fuzzy approach is that it is artificial or implausible or somehow inappropriate to associate each vague *predicate* in natural language with a particular function that assigns a real number to each object (the object's degree of possession of the property picked out by that predicate) and to associate each vague *sentence* in natural language with one particular real number (the sentence's degree of truth).

I have addressed this problem in detail in Smith (2008, 2011). I have argued that the problem is a genuine one and that at its core are considerations about how meanings are determined. It is generally accepted that language is a human artefact. The sounds we make mean what they do because of the kinds of situations in which we (and earlier speakers) have made (and would make) those sounds (e.g. had the word 'dog' always been used where 'cat' was used, and vice versa, then 'dog' would have meant what 'cat' in fact means, and vice versa). So there is an essential connection between *meaning* and *use*. More precisely, consider the following kinds of facts:

- All the facts as to what speakers actually say and write, including the circumstances in which these things are said and written, and any causal relations obtaining between speakers and their environments.
- All the facts as to what speakers are disposed to say and write in all kinds of possible circumstances.

There is widespread agreement in the literature that semantic facts are never primitive or brute—they are always determined by some other, meaning-determining facts. There is also widespread agreement that the meaning-determining facts are the ones just itemised: facts about actual usage, counterfactual usage and usage dispositions.[1] Thus, if these facts are insufficient to determine (unique) meanings for some utterances, then those utterances have no (unique) meanings.

This generates a constraint on any theory of vagueness: if the theory says that vague predicates have meanings of such-and-such a kind (e.g. fuzzy sets), then we must be able to satisfy ourselves that the meaning-determining facts could indeed determine such meanings for actual vague predicates. To the extent that the meaning-determining facts do *not* appear sufficient to determine meanings for vague predicates of the kind posited by some theory of vagueness, that theory is undermined.

This is precisely where the basic fuzzy theory of vagueness—according to which each vague discourse is associated with a unique intended fuzzy model—runs into trouble. For it certainly seems that usage and usage dispositions do not suffice to pick out a particular function from objects to fuzzy truth values representing the

[1] Some authors would also like to include other kinds of facts, such as facts concerning the eligibility as referents of objects and sets, or facts concerning the simplicity/complexity of interpretations. Even on these views, however, usage plays a central role.

extension of 'is tall' (and similarly for other vague predicates). Our tendencies to classify certain people as (clearly) tall, certain others as (clearly) not tall, and to hedge over other cases (the borderline cases)—including any tendencies we might have to be more or less confident or reticent about different borderline cases—just do not, it seems, form a rich enough base to determine a *unique* function that assigns a degree of tallness (a real number between 0 and 1) to each object. Keefe asks (in the passage quoted above) "what could determine which is the correct function, settling that my coat is red to degree 0.322 rather than 0.321"? The answer, I have argued, is that what could determine this is our usage (actual and counterfactual/dispositional). But now the problem: our usage *does not* (it would seem) determine a unique function. Hence the basic fuzzy theory of vagueness is the wrong model of vague discourse.

The story does not end there: I have also presented and argued for a solution to the problem. The solution is to move from a basic fuzzy theory of vagueness (on which each vague discourse is associated with a unique intended fuzzy model) to a theory that I call *fuzzy plurivaluationism* (on which each vague discourse is associated with multiple acceptable fuzzy models). But these further developments are not relevant for our present purposes. We shall turn now to a problem for Bayesian epistemology that is, on the surface, very similar to the artificial precision problem for the basic fuzzy theory of vagueness.

3 Bayesian Epistemology

'Bayesian epistemology' denotes a family of views sharing three core tenets:

1. An agent's 'belief state' can be represented by an assignment to propositions of real numbers between 0 and 1 inclusive.[2]
2. In so far as the agent is rational, this assignment will obey the laws of probability.[3]
3. In so far as the agent is rational, she will update her assignment in the face of new evidence by conditionalisation.[4, 5]

[2] Different Bayesians use different terms where I have used 'belief state'—e.g. 'credal state', 'credences', 'degrees of belief', 'degrees of confidence', and 'subjective probabilities'. At this point I use the term 'belief state' with scare quotes and without an explanation of what the term means, because a key issue below—beginning in Sect. 4—will be what, exactly, the agent's 'belief state' is supposed to be.

[3] Taking propositions to be sets of possible worlds, the second tenet becomes the requirement that the assignment of real numbers to propositions constitutes a probability measure over the space of possible worlds.

[4] Given tenet 2, the agent's assignment of real numbers to propositions can be referred to as a probability assignment. Tenet 3 then requires that the agent's posterior probability assignment to a proposition P (after the evidence E comes in) is equal to her prior (before the evidence comes in) conditional probability of P given E.

[5] Some characterisations of Bayesianism (e.g. Joyce 2010, 282) add a fourth tenet: rational agents make decisions by maximising expected utility.

Tenet 1 is a descriptive claim and tenets 2 and 3 are normative claims. Bayesian epistemology is thus partially descriptive and partially normative. Tenet 1 says that agents can in fact be represented as assigning numbers to propositions (a descriptive claim), and tenet 2 says that these assignments should (although in fact they may not always) obey the laws of probability (a normative claim; and similarly for tenet 3).

In this paper my main focus is on tenet 1—and in particular, on a certain kind of objection to it.[6] The objection is that—in supposing that agents can be represented by a unique assignment of real numbers to propositions—it involves excessive or artificial precision:

> I find it wildly implausible to suppose (as some orthodox Bayesians have done) that actual investigators in general harbor precise, real-valued degrees of confidence for hypotheses. Even when construed as a regulative ideal, the requirement that investigators harbor such precise degrees of confidence looks as if it owes more to an unfortunate worship of false precision than it does to reason (Kaplan 1996, x).

> the orthodox Bayesian account of belief is inaccurate. The nature of the inaccuracy is that it requires point-valued credences, which carries a commitment to the existence of sharp lines where, intuitively, there aren't any (Rinard 2017, 262).

> In general, our opinion is certainly not expressible in precise numerical probabilities (Van Fraassen 1990, 345).

The objection is that it is artificial or implausible or somehow inappropriate to represent an agent's belief state by an assignment of unique numerical values to each proposition.

Why? Is there an argument here? There are indeed arguments in the literature. Before we look at these existing arguments in Sect. 6, however, I want (in Sect. 5) to explore whether there is a good argument to this conclusion that runs along similar lines to the dispositional argument against precision in the case of fuzzy theories of vagueness—and before undertaking this exploration, I shall (in Sect. 4) introduce an important distinction.

4 Two Aspects of Thinking Agents

There is an influential view, going back to Ramsey, according to which degrees of belief are strengths of behavioural dispositions:

> the degree of a belief is a causal property of it, which we can express vaguely as the extent to which we are prepared to act on it. …it is not asserted that a belief is an idea which does actually lead to action, but one which would lead to action in suitable circumstances…The difference [between believing more firmly and believing less firmly] seems to me to lie in how far we should act on these beliefs (Ramsey 1926, 65–6).

On this view, one's degree of belief that S is a measure of the strength of one's tendency to act as if S.

---

[6]Objections have also been made to tenets 2 and 3—but they are not my concern in this paper.

There is another widespread view of degrees of belief, according to which an agent's degrees of belief constitute her view about how the world is, as justified by the evidence available to her. For example, (Walley 1991, 14) writes that "Epistemic probabilities…depend on the available evidence" and (Sturgeon 2008, 159) writes of "evidence and attitude aptly based on it."

On the first conception—which I'll call the *dispositional* conception—degrees of belief are essentially related to (possible) action: they are (strengths of) behavioural dispositions. On the second conception—which I'll call the *epistemic* conception—degrees of belief are essentially related to evidence and rational judgement: they are evidence-based opinions about the way the world is.

Most authors seem simply to adopt one conception without mentioning or acknowledging the other. Some authors do note the distinction—but even then there is a tendency to write as if what we have here are two rival conceptions of one thing ('degrees of belief'). I think it is much more fruitful to view this distinction—at least initially—as a distinction between two aspects of thinking agents. Both aspects are present—and we need accounts of both.

On the one hand, thinking agents need to keep a running view of the way the world is, and this view needs to be appropriately determined or justified by the evidence available to them. There should not be anything in this picture of the world that isn't warranted by the available evidence. Now sometimes the evidence tells us clearly that something is or is not the case, but at other times the evidence is less decisive—and so we should suppose that the agent's running view of how the world is will be a matter of degree rather than all-or-nothing: some ways the world might be are definitely ruled out as not being the case; some are definitely ruled in; but others are only more or less ruled out or in.

On the other hand, thinking agents act in the world—they make choices and decisions—and at any time, they have dispositions to make these choices in certain ways. Knowing how an agent will choose can be crucial to interacting with him. These dispositions are also a matter of degree, rather than all-or-nothing. An agent need not have a full-on tendency to make just one choice in a certain kind of situation—and no tendency at all to make any of the other choices that would also be possible in that situation. Rather, he may have stronger or weaker tendencies to make various choices.

Clearly both aspects are, and should be, present in typical thinking agents. There is no reason to think we can get a full understanding if we focus on one aspect alone. Nor should we assume, in advance, that these two aspects of thinking agents must really, at bottom, be the same aspect. There's certainly reason to think there should be some relation between them—any view will need to allow that the choices of *rational* agents, at least, should be sensitive to their view of how the world is—but just how close the relationship should be is a big question. I cannot explore this question in detail here—but I can at least note some possible positions. Certain dispositionalist or functionalist views of belief would seem to suggest that ultimately there can be little or even no real distinction here: that having (say) a mid-strength tendency to act as if P is necessary and sufficient for the attribution of a mid-strength epistemic degree of belief that P. Other views of belief—for example certain representationalist views—would seem to allow room for a real distinction between an evidence-based and rationally inferential realm of epistemic beliefs and a realm of behavioural dispositions. Yet other views would seem to demand a significant distinction here: see for example Gendler (2008a, b) on alief and belief. In the end, then, it *may* be that there is a good theoretical argument leading to one or more of the following conclusions: that the epistemic and dispositional aspects of thinking agents are ultimately one and the same aspect; that an agent's epistemic degree of belief that *S*—the degree to which *S*'s being the case forms part of her evidence-based picture of how the world is—should always equal her dispositional degree of belief that *S*—the strength of her tendency to act as if *S*; or that there is some tool—for example probability theory—that models both aspects of thinking agents in one go. But given the deep theoretical issues involved, we should not *begin* by assuming that any of these conclusions are true. We should start by acknowledging two aspects of thinking agents—epistemic and dispositional—and the need for theories of both of them, even if they might turn out in the end to be the same theory. Certainly we should not start with a theory—for example probability theory—and assume that it is the only possible tool available, or that we can use it to model everything.

While it is not common, the viewpoint that I have just outlined—according to which we have a distinction between two coexisting aspects of thinking agents (dispositional degrees of belief and epistemic degrees of belief), rather than two rival conceptions of one aspect ('degrees of belief')—is shared by some other authors. Kyburg Jr. (1983) is a possible example—and the Transferable Belief Model (Smets and Kennes 1994) is a clear example of a well worked-out framework in which this distinction plays a crucial role.[7]

5 A Dispositional Argument Against Bayesianism?

The artificial precision problem for the basic fuzzy theory of vagueness ultimately boils down to usage dispositions: those dispositions are not rich or detailed enough to determine unique fuzzy sets as meanings for vague predicates. The question now is whether there is a genuine problem *of the same sort* for Bayesian epistemology: that is, a problem to do with our dispositions not being rich or detailed enough to determine unique numerical assignments to propositions. (In the Bayesian case, these numerical assignments represent degrees of belief, not degrees of truth.) Note that if there is such a problem, it is primarily a problem for tenet 1 of Bayesianism.

In the previous section I distinguished two notions of degree of belief: degree of belief that *S* as strength of tendency to act as if *S* (dispositional degree of belief); and degree of belief that *S* as firmness of evidence-based opinion as to whether *S* is true (epistemic degree of belief). When it comes to dispositional arguments, the

[7]The epistemic/dispositional distinction that I have just drawn is not the same as the traditional distinction between occurrent and dispositional belief: the latter is most naturally viewed as a distinction amongst epistemic beliefs.

action must centre on dispositional degrees of belief. Degrees of belief of this kind just are (strengths of) behavioural dispositions—so to the extent that an agent lacks behavioural dispositions, it ipso facto lacks (determinate) degrees of belief of this kind. Epistemic degrees of belief, on the other hand, either are not subject to a dispositional argument at all, or are subject to such an argument only via a dispositional argument concerning dispositional degrees of belief. Which of these possibilities one takes to be the case will depend on one's substantive view of the relationship between dispositional and epistemic degrees of belief. On a view on which epistemic degrees of belief are to a large extent independent of behavioural dispositions, there can be no dispositional argument against epistemic degrees of belief. On this sort of view, the notion of epistemic degree of belief—strength of opinion as justified by the available evidence—is not a dispositional notion. A purely contemplative being with no behavioural dispositions—for example a computer programme—might have degrees of belief in this sense. Lack of dispositions to do things is no threat to the existence or determinacy of degrees of belief of this kind. On a view on which epistemic degrees of belief are more closely tied to behavioural dispositions—for example a view on which having (say) a mid-strength tendency to act as if P is necessary and sufficient for the attribution of a mid-strength epistemic degree of belief that P—there will be room for a dispositional argument against determinate epistemic degrees of belief, but it will go via a dispositional argument against determinate dispositional degrees of belief. On this sort of view, a dispositional argument can get a grip on epistemic degrees of belief only by first getting a direct grip on dispositional degrees of belief and then transferring that grip via theoretical claims linking (or even identifying) dispositional and epistemic degrees of belief. Thus, either way, when it comes to dispositional arguments, the action must centre on dispositional degrees of belief.

So let's work with the conception of degree of belief as strength of tendency to act, and investigate whether a dispositional argument shows that our degrees of belief (of this kind) cannot be represented by unique numerical assignments. Let A be a typical human agent. The argument would be that for many or all propositions S, A's behavioural dispositions are insufficiently rich to enable the attachment of a unique numerical degree to A's tendency to act as if S. Now considered in the abstract, this may seem plausible. But thanks to de Finetti and others, we can see that it is incorrect. Consider an agent A and an amount of money that is neither so small that A would not care about winning or losing such an amount nor so big that A would not want to bet such an amount if there were any risk at all of losing the bet. For the sake of argument, let the amount be \$1. Now take a proposition S. There is to be a bet on S. There are two parties to the bet. One party will take on the Pro-role: will win the bet if S turns out to be true (and lose if S turns out to be false). The other party will take on the Con-role: will win the bet if S turns out to be false (and lose if S turns out to be true). The *betting ratio* x of the bet determines how much is won and lost: x is a real number between 0 and 1 (inclusive) such that Pro pays $\$x$ to bet, Con pays $\$(1-x)$ to bet, and the winner takes all (i.e. $\$x + \$(1-x) = \$1$). Now A is to set the betting ratio x in such a way that she would be happy to play the Pro role or the Con role. (Compare cutting a cake into two pieces in such a way that you would be happy to get either of them.) The idea is that x can then be taken as a numerical representation of the strength of A's tendency to act as if S.

Now it seems to me that for most agents A and most propositions S, if A were in a situation where a bet on S is to take place and A must set the betting ratio x, A would set some x. Of course there might be some cases in which A freezes up completely and is simply unable to set x—but I think that would not be the norm. If one really were required to set x, one would. Hence one *does* have dispositions to set x. So the dispositional argument fails: our behavioural dispositions *are* rich enough to underpin unique numerical assignments to degrees of belief (conceived as strengths of tendency to act).

Several issues need to be clarified at this point.

(1) Do not confuse sitting in an armchair thinking about an imaginary betting situation with really being in a situation where you must set the betting ratio. Maybe you have no idea what x you'd set when you imagine the situation. Maybe you can't imagine setting x at all. This is beside the point. My claim is that you *would* set x and hence you *do* have a disposition to set x.

(2) I have taken from de Finetti the idea of measuring degrees of belief by betting ratios—but I am not adopting the operationalist perspective that he expresses in passages such as the following:

> In order to give an effective meaning to a notion—and not merely an appearance of such in a metaphysical-verbalistic sense—an operational definition is required. By this we mean a definition based on a criterion which allows us to measure it. We will therefore be concerned with giving an operational definition to the prevision of a random quantity, and hence to the probability of an event (de Finetti 1974, 76).

The picture I'm working with is this. An agent, in this actual world, has a determinate strength of tendency to act as if S. This is a dispositional property of the agent, grounded in facts about his brain and body. Betting behaviour is one, but *not* the only kind of behaviour that falls under the heading 'acting as if S': you can act as if it will rain by betting that it will but also by taking your umbrella. The agent's strength of tendency to act as if S certainly isn't *defined* as the number he would give in another, nearby world if he had to set a betting ratio (in that world). However, thinking about such a world enables us to see that the agent does indeed have rich and detailed dispositions: rich and detailed enough to warrant modelling his strength of tendency to act as if S as a particular real number. The situation is analogous to the case of length measurement. I suppose that the table before me has some determinate length. I do not *define* this length as the number I would get if I measured the table with a tape measure. But if someone were to doubt that there are any facts about the table that warrant modelling its length as a particular real number, we could point out that were we to lay a tape measure on top of the table, we would see a particular mark on the tape measure lined up with the edge of the table, and there would be a particular number associated with that mark. Even if we never measure the table, there are facts about it that ensure we would get a particular answer if we did—and this justifies modelling its length as a particular number (as opposed, say, to a coarse-grained model in which we model lengths as 'short', 'medium', and 'long').

Note that this line of thought is supposed to function only as a ‘proof of concept’—not as a specification of a standardised measurement system. If we want the *official* length of the table—say for some legal purpose, such as determining whether a contract has been breached—it has to be measured by a trained professional using certified measuring equipment. Similarly, if we wanted an official measurement of Bob’s degree of belief that *S*, we would need to settle on a particular formulation of the bet to be offered, we would need to train and certify people to offer such bets in a particular way, and so on. Presumably a body such as the Australian National Measurement Institute could settle such matters—no doubt with some measure of arbitrariness and stipulation—if there were a need or reason to do so. But my purpose here is not to specify a particular procedure by means of which we can actually obtain a unique number as the measure of Bob’s degree of belief that *S*: it is simply to make it plausible—by thinking about how agents would set betting ratios—that their dispositions are sufficiently rich and detailed to warrant modelling strengths of tendency to act as particular numbers.

With these points in mind, we can respond to several potential worries.

(2a) Do not object that de Finetti’s procedure fails to get us sufficient precision: that it could never justify assigning, say, 0.3 rather than 0.300001 as the measure of some tendency to act. The fact that we can only measure lengths to a certain level of precision—say, down to the nearest millimetre if we are using an ordinary tape measure—does not threaten the idea that objects have unique lengths which can be represented by real numbers. Limited precision is a given in all cases of numerical measurement: it is not a special problem for degrees of belief.

(2b) Consider an agent *S* in a particular context *C*, in which *S* is not in fact asked to set a betting ratio for a bet on a proposition *P*. I have said that if *S* were asked to set a betting ratio, she would. The worry now is that if we think about this in the way advocated by Lewis (1973), there might not be a single determinate ratio that *S* would set: it may be that in some of the worlds most similar to the actual world, *S* says 0.29, while in others—in which the question is asked in a slightly different tone of voice, or by a different questioner—*S* says 0.3. One response to this worry is to say that we should not think about the situation in Lewis’s way, but rather in Stalnaker’s way, according to which there is a unique most similar world in which the agent is asked to set a betting ratio (Stalnaker 1968). A second response is to say that even if we stick with Lewis’s conception, the resulting indeterminacy is not a special problem for degrees of belief, but is of a piece with indeterminacy in other cases of measurement. For example, if the in-fact-unmeasured table were measured, the resulting length might be slightly different depending on who performed the measurement and what kind of measuring device was used. If we wanted an official length we would need to have a standardised measuring system and accredited measurers—and similarly for betting ratios. There is nothing here to cast doubt on the thought that the table has a determinate length: it is precisely because the table is the determinate way it is that we would get *this* answer if we measured it in this way and *that* answer if we measured it in that way. Similarly, there is nothing here to cast doubt on the thought that the agent has a determinate strength of tendency to act as if *S*: it is precisely because the agent is the determinate way she is that she would set *this* ratio if asked in this way and *that* ratio if asked in that way.

(3) Someone might accept my core point that we *would* set certain betting ratios and hence *do* have dispositions to set them, but question the relevance of this to degrees of belief. That is, why should we take these betting ratios as degrees of belief? Well, because we are, at the moment, taking degree of belief to be strength of tendency to act—and it just seems undeniable that setting a higher betting ratio in a bet on *S just is* (one way of) acting more strongly as if *S* is true. To see the point more clearly, consider again the case of fuzzy theories of vagueness. Suppose I were to give you a corkboard and a pin. I then present you with pairs of objects and predicates. For each pair, you must stick the pin in the corkboard. (These are the only instructions I give you.) You would do so—and hence you have dispositions to do so. Now suppose I propose to measure the distance from the left side of the corkboard to the pin (in a system of units where the width of the corkboard is 1 unit) and take that to be a measure of the degree to which the object falls under the predicate. In this case, it seems clear that although the dispositions are present, they are irrelevant: there seems to be no good reason why we should take these dispositions to determine *degrees of truth* (or degrees of application of predicates, or degrees of membership in fuzzy sets). If we now return to the case of degrees of belief—conceived as strengths of tendency to act—I think it should be clear that there is no such problem: evidently, setting a higher betting ratio in a bet on *S just is* (one way of) acting more strongly as if *S* is true.

(4) It seems unlikely that the betting ratios that a typical actual agent would set will satisfy the probability axioms. But note that—if correct—this observation does not speak against tenet 1 of Bayesianism. Nor, moreover, does it speak against tenet 2: for tenet 2 of the Bayesian approach is normative. It holds, not that an agent's degrees of belief *do* satisfy the laws of probability, but that they *should*—that in so far as they do not, the agent is irrational. And even if actual agents do not satisfy this norm, there seems to be a strong reason why they should—that is, a strong reason to accept it as a norm. The reason is that if an agent's betting ratios do not satisfy the probability axioms, then the agent is susceptible to Dutch book—to a sure loss. Now some have objected to Dutch book arguments on the grounds that they are pragmatic and cannot justify claims about epistemic rationality. These objections may or may not be well founded when directed against Dutch book arguments to the conclusion that *epistemic* degrees of belief should obey the laws of probability. However for now we are talking about *dispositional* degrees of belief—strengths of tendency to act—and it seems quite clearly irrational to have tendencies that will leave you worse off *no matter what*. When we are thinking of degrees of belief in this practical sense—as tendencies to action—a pragmatic argument is just the sort of argument we need.

(5) It seems quite possible that the betting ratios that a typical actual agent would set will be variable, unstable, and sensitive to context. That is, given the very same proposition *S*, an agent might set a different ratio for a bet on *S* if asked to do so in

context C_1 from that she would set in context C_2.[8, 9] Note again, however, that—if correct—this observation does not speak against tenet 1 of Bayesianism. It means that an agent's behavioural dispositions might change with time and with context. This does not undermine my key claim that—at any given time and in any given context—the agent's behavioural dispositions are sufficiently determinate to enable the strength of her tendency to act as if S to be measured by a particular real number. Of course, if an agent's strength of tendency to act as if S changes between t_1 and t_2, even though the agent receives no new evidence during that period that bears one way or the other on whether S is true, then further issues do arise. One issue is that the agent will have violated tenet 3 of Bayesianism: she will not have updated by conditionalisation. However, as in the case of tenet 2, tenet 3 is normative: it holds not that agents *do* always update by conditionalisation but that they *should*. The observation that ordinary folk do not conform to tenet 3 does not automatically mean that we should abandon it as a normative principle. (Should we abandon it? I think the answer is less clear here than in the case of tenet 2. There is a diachronic Dutch book argument for taking tenet 3 as a normative principle—but this argument seems less secure than the synchronic Dutch book argument for tenet 2. For example, in the Transferable Belief Model, updates need not respect conditionalisation—yet Smets (1993) argues that this model is *not* refuted by the diachronic Dutch book argument.) A second issue is that, even setting aside the particular question of conditionalisation, it might seem that the agent has done something *irrational* in changing her degree of belief that S when no relevant evidence has come in: for shouldn't degrees of belief be determined by the available evidence? Well, *epistemic* degrees of belief should be—but with *dispositional* degrees of belief (which are our current concern), change over time is surely to be expected, with or without change in evidence: degrees of belief, on this conception, are strengths of tendencies to act—and of course your dispositions may change over time, for all sorts of reasons.

(6) Someone might object that while it is true that an agent *would* set a particular betting ratio, this does not mean that she has any *disposition* to do so. Degree of belief, on the current conception, is supposed to be strength of *tendency* to act.

[8]Note that this is a different kind of variation from the one we considered under (2b) above. There we were talking about a certain agent S in a certain context C, and imagining how S would set a betting ratio if asked to do so in C. Given that S is not actually asked in C, this means considering nearby worlds in which S is asked—and the thought was that S might not set the very same ratio in all of these worlds. Now the situation is different. We are imagining S first in actual context C_1 and then in actual context C_2, and we are imagining how S would set a betting ratio if asked to do so in each of these contexts. So now we are considering the ratio S sets in a counterfactual situation similar to C_1 and the ratio S sets in a counterfactual situation similar to C_2, where C_1 and C_2 are different contexts in the actual world (as opposed to the ratios S sets in two counterfactual situations, both of which are similar to the same actual context C).

[9]There is psychological evidence that agents' explicit judgements or estimates of probability are variable and situation-dependent. This does not automatically mean, however, that agents' behavioural dispositions are variable and context-dependent—that would follow only if degrees of belief (in the dispositional sense) are straightforwardly mirrored in estimates of probability. The *possibility* that I am considering now is that agents' behavioural dispositions *are* variable and context-dependent.

The objection now is that the agent *has no* tendency to act. She would act, but this act would not be determined by any underlying tendency or disposition: it would be an undetermined act ex nihilo that she would come up with entirely on the fly. In response, I don't really have anything to say except that I find this completely implausible, as a general picture. Our actions may not always be rational or justified, but in general it is natural to suppose—at least in the absence of some compelling argument or evidence to the contrary—that they stem from underlying states of our brains and bodies that cause us to act in the ways we do. The objection under consideration is that when asked in some situation to specify a betting ratio on S, I might specify 0.3 (say)—but I could *just as well* have specified 0.4 or some other number (in the same situation): the number I specify pops out of nowhere, rather than being a manifestation of an underlying disposition or tendency to specify that number. Now here we must be careful to distinguish the two kinds of degree of belief. On the *epistemic* conception it may be true to say that 0.3 'comes from nowhere' in the sense that this number is not uniquely *justified* by the evidence available to me. But on the *dispositional* conception (which is our current concern), the claim that my answer of 0.3 'comes from nowhere' in the sense of popping out randomly—in such a way that other numbers could just as easily have popped out, and it simply happened for no underlying cause to be this number—seems to me utterly far-fetched.

6 Arguments in the Literature

In the previous section I argued that there is no good argument against tenet 1 of Bayesianism along the lines of the dispositional argument against the basic fuzzy theory of vagueness. In this section I turn to actual arguments in the literature that object to Bayesian epistemology on the grounds that it involves false precision. Some of these arguments can be seen as (more or less detailed gestures in the direction of) dispositional arguments—and can be seen to fail for reasons already discussed. Some of them can be seen as being (to a greater or lesser extent) in the same ballpark as the dispositional argument—and can also be seen to fail. Finally, some of them can be seen as entirely different kinds of argument. With respect to this last group, my aim here is to point out the differences (from dispositional arguments)—not to go on to assess whether these other arguments are sound.

Joyce (2010, 283) writes:

> numerically sharp degrees of belief are psychologically unrealistic. It is rare, outside casinos, to find opinions that are anywhere near definite or univocal enough to admit of quantification. An agent with a precise credence for, say, the proposition that it will rain in Detroit next July 4th should be able to assign an exact 'fair price' to a wager that pays $100 if the proposition is true and costs $50 if it is false. The best most people can do, however, is to specify some vague range.

Joyce seems to be claiming that in general people could not and would not set a betting ratio if required to do so. This just strikes me as false. As discussed in Sect. 5, for most agents A and most propositions S, if A were in a situation where a bet on

S is to take place and A must set the betting ratio x, A would set some x. Of course there might be some cases in which A freezes up completely and is simply unable to set x—and of course if we simply *asked* A what x she would set if she were required to (as opposed to requiring her to set x), A might not be able to give an answer (recall point (1) in Sect. 5)—but in general I think that if one were required to do so, one would set some x.

Walley (1991) writes:

> We know, through introspection, that our beliefs about many matters are indeterminate…Imprecise probabilities are needed to model the indeterminacy (4).
>
> It seems clear that indeterminacy exists. A little introspection should suffice to convince You that Your beliefs about many matters are presently indeterminate (210).

There are different ways of interpreting this. Recall the distinction in Sect. 4 between the dispositional conception of degrees of belief as strengths of tendency to act and the epistemic conception of degrees of belief as evidence-justified view of how the world is. Walley might be thinking of beliefs in the epistemic sense. Now perhaps it is plausible that our epistemic degrees of belief are accessible to introspection: they are, after all, supposed to constitute our considered, evidence-justified opinions about the way the world is. Interpreted this way, the argument is very different from a dispositional argument. Unlike a dispositional argument, it is directly concerned with epistemic degrees of belief, not dispositional degrees of belief. It forms a descriptive counterpart (introspection tells us that epistemic degrees of belief *are not* determinate) to the normative argument that we shall see below (epistemic degrees of belief *should not* be determinate). For our present purposes it is not relevant to assess these arguments: pointing out that they are of a completely different nature from the dispositional argument against the basic fuzzy theory of vagueness is enough.

So suppose we interpret Walley as talking about dispositional degrees of belief. Now if we are concerned with degrees of belief as strengths of tendency to act then there is no reason to think that introspection should have the final word. That is, our dispositions might be far more determinate than we can ascertain by introspection. For, in general, there is no reason to think that we can always imagine a situation in sufficient detail to trigger (in imagination) our dispositions and hence allow us to know how we would react in that situation. You might have no idea how you would react in a certain dangerous situation—or you might think that you would run away or freeze up completely. Nevertheless it might still be the case that you would act courageously in that situation, were you to face it. Similarly, as we have already discussed, if you imagine a betting situation, you might not have much idea what ratio you'd set—or you might think that you would not set one at all. Nevertheless, were you in such a situation and required to set a betting ratio, you would (in general, I maintain) do so.

Walley (1991) writes:

> If You are required to choose between X and Y then You will, of course, choose one way or the other, but in cases of indeterminacy Your choice is simply not determined by Your current state of mind. Your mind is not 'made up' (209–10).

> It is essential to distinguish between choice and preference. A **choice** is a decision about how to act, that is made in a specific context…A **preference** is an underlying disposition to choose in a particular way…You can choose a_1 over a_2 without having any preference between them. A choice can be **arbitrary**, in the sense that it is not determined by Your preferences, beliefs and values (236–7).[10]

Again, there are different ways of interpreting this. We could interpret Walley as expressing the idea discussed under point (6) in Sect. 5: even if there is determinate counterfactual behaviour, that does not mean that there are determinate behavioural *dispositions*. On the contrary, the behaviour 'just happens'—rather than being a manifestation of an underlying disposition to behave in that way. I have already responded (in Sect. 5) to this line of thought. A different interpretation is that sometimes we act in ways that we cannot justify on the basis of our underlying *epistemic* beliefs (and preferences and values). It's not that our actions 'just happen' in the sense just mentioned: on the contrary, at any time, our brains and bodies are in such a state that we have definite behavioural dispositions. Rather, the point is that sometimes (or often) our actions are arbitrary in the sense that we have no more reason or justification for acting in the way we do than in some other possible way. Now this may be true—but it is a very different kind of point from the dispositional argument against the view that our dispositional degrees of belief can be represented by particular numbers.

Following the passage quoted near the beginning of this section, Joyce (2010) continues:

> While psychological implausibility is one worry, a more decisive problem is that precise degrees of belief are the wrong response to the sorts of evidence that we typically receive. As argued in Joyce (2005), since the data we receive is often incomplete, imprecise or equivocal, the epistemically *right* response is often to have opinions that are similarly incomplete, imprecise or equivocal (283).
>
> Precise credences…always commit a believer to extremely definite beliefs about repeated events…even when the evidence comes nowhere close to warranting such beliefs (285).

A similar line of thought has been expressed by many other authors—for example Kaplan (1996):

> The moral would seem to be that if we want to give evidence its due—if we want to maintain that a rational investigator ought to adopt only those states of opinion she has good reason to adopt—we had better conclude that Immodest Connectedness is not a principle you should want to satisfy (27–8).
>
> It is important to notice that my reason for rejecting as falsely precise Immodest Connectedness's demand that you place a monetary value on each well-mannered state of affairs is not what one might have expected. It is not that this demand is not humanly satisfiable. For if *that* were all that was wrong, the demand might still play a useful role as a regulative ideal—an ideal which might then be legitimately invoked to get you to 'solve' your decision problem as the orthodox Bayesian would have you do. My complaint about the orthodox Bayesian demand is rather that it imposes the wrong regulative ideal. For if you place a monetary value on each well-mannered state of affairs, you have a determinate assignment of *con*(–) to every hypothesis—and then you are not giving evidence its due (29).[11]

[10] Cf. also Walley (1991, 106, 245, 247, 533) and Eriksson and Hájek (2007, 189–90).

[11] Cf. also Kaplan (1996, 24.)

Walley (1991):

> When there is little relevant evidence, even the 'ideal' probabilities are imprecise (7).
>
> When there is little or no relevant evidence, the probability model should be highly imprecise or vacuous. More generally, the precision of probability models should match the amount of information on which they are based (34).
>
> 'Vagueness' or lack of information should be reflected in imprecise probabilities (246)

and Levi (1985, 396):

> it is sometimes rational to make no determinate probability judgment and, indeed, to make maximally indeterminate judgments. Here I am supposing…that refusal to make a determinate probability judgment does not derive from a lack of clarity about one's credal state. To the contrary, it may derive from a very clear and cool judgment that on the basis of the available evidence, making a numerically determinate judgment would be unwarranted and arbitrary.[12]

The argument is that degrees of belief should be determined and justified by the available evidence; but often the available evidence is sparse or vague; and in such cases, degrees of belief should not be precise. This may, or may not, be a powerful line of argument—but it suffices for present purposes to point out that it is clearly an argument that degrees of belief in the epistemic (rather than dispositional) sense should be imprecise—and it is clearly a very different kind of argument from the dispositional argument.

7 Conclusion

The artificial precision argument against fuzzy theories of vagueness and the false precision objection to Bayesian epistemology have the same surface form. A more detailed investigation, however, has revealed deep differences. The kind of dispositional argument that carries weight in the fuzzy case does not, I have argued, go through in the Bayesian case. The main purpose of this investigation has been to shed light on the conceptual landscape. But it may be that there is a more substantive conclusion in the offing. Suppose that the Bayesian account is correct for dispositional degrees of belief. (Note that I have not shown this: I have defended the Bayesian view—and in particular, its first tenet—against a certain objection; I have not given a complete defence of the full Bayesian account.) Suppose also that the final kind of objection considered in Sect. 6 is correct—in which case the Bayesian theory is inadequate as an account of epistemic degrees of belief. (Note that this goes far beyond anything argued in this paper—but just suppose for the moment that it is true, in order to see what follows.) In that case, we need different accounts of dispositional degrees of belief and epistemic degrees of belief. That sort of two-level picture is not unheard of—it is one of the core ideas of the Transferable Belief Model (Smets and

[12] Levi here refers to Peirce, Fisher, Neyman, Pearson and Kyburg as examples of earlier authors who held similar anti-Bayesian views. Cf. also Levi (1974, 394–5.)

Kennes 1994)—but it is not common: most of the debate in this area is concerned with finding the right model of 'degrees of belief' (or 'credences', or whatever other term is chosen). The more substantive conclusion that may be in the offing is that what we really need are two different models (and an account of their interactions): one for dispositional degrees of belief and one for epistemic degrees of belief.[13]

References

de Finetti, B. (1974). *Theory of probability: A critical introductory treatment* (Vol. 1). Chichester: Wiley.

Eriksson, L., & Hájek, A. (2007). What are degrees of belief? *Studia Logica*, *86*(2), 183–213.

Gendler, T. S. (2008a). Alief and belief. *Journal of Philosophy*, *105*, 634–663.

Gendler, T. S. (2008b). Alief in action (and reaction). *Mind and Language*, *23*, 552–585.

Haack, S. (1979). Do we need fuzzy logic? *International Journal of Man-Machine Studies*, *11*, 437–445.

Joyce, J. M. (2005). How degrees of belief reflect evidence. *Philosophical Perspectives*, *19*, 153–179.

Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, *24*, 281–323.

Kaplan, M. (1996). *Decision theory as philosophy*. Cambridge: Cambridge University Press.

Keefe, R. (1998). Vagueness by numbers. *Mind*, *107*, 565–579.

Kyburg, H. E., Jr. (1983). Two world views. *Epistemology and inference* (pp. 18–27). Minneapolis: University of Minnesota Press.

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, *71*(13), 391–418.

Levi, I. (1985). Imprecision and indeterminacy in probability judgment. *Philosophy of Science*, *52*(3), 390–409.

Lewis, D. (1973). *Counterfactuals*. Cambridge: Harvard University Press.

Ramsey, F. P. (1926). Truth and probability. Ramsey (1990) (pp. 52–94).

Ramsey, F. P. (1990). D. H. Mellor (Ed.), *Philosophical papers*. Cambridge: Cambridge University Press.

Rinard, S. (2017). Imprecise probability and higher order vagueness. *Res Philosophica*, *94*, 257–273.

Smets, P. (1993). No Dutch book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. In R. Scozzafava (Ed.), *Workshop on probabilistic expert systems* (pp. 181–204). Rome: Societa Italiana di Statistica.

Smets, P., & Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, *66*, 191–234.

Smith, N. J. J. (2008). *Vagueness and degrees of truth*. Oxford: Oxford University Press.

Smith, N. J. J. (2011). Fuzzy logic and higher-order vagueness. In P. Cintula, C. G. Fermüller, L. Godo, & P. Hájek (Eds.), *Understanding vagueness: Logical, philosophical and linguistic perspectives* (pp. 1–19). Studies in logic, Vol. 36. College Publications.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford: Basil Blackwell.

Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, *42*(1), 139–165.

Urquhart, A. (1986). Many-valued logic. In D. Gabbay, & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. III, pp. 71–116). Dordrecht: D. Reidel.

[13]Thanks to Richard Dietz, Alan Hájek, Brian Hedden, Dan Lassiter, an anonymous referee, and audiences at a Current Projects seminar at the University of Sydney on 23 March 2017 and a Department of Philosophy seminar at the University of Adelaide on 7 September 2018 for helpful comments.

Van Fraassen, B. C. (1990). *Figures in a probability landscape* (pp. 345–356). Dordrecht: Kluwer Academic Publishers.
Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall.

Regret, Sub-optimality, and Vagueness

Chrisoula Andreou

Abstract This paper concerns regret, where regretting is to be understood, roughly, as mourning the loss of a forgone good. My ultimate aim is to add a new dimension to existing debate concerning the internal logic of regret by revealing the significance of certain sorts of cases—including, most interestingly, certain down-to-earth cases involving vague goals—in relation to the possibility of regret in continued endorsement cases. Intuitively, it might seem like, in continued endorsement cases, an agent's regret (if it is to make sense) must be tied to the idea that the forgone good is no better than the achieved good but is also not fully made up for by the achieved good because the goods are (too) different in kind. But this view is controversial. After describing a challenge to the view, as well as the main features of the debate regarding regret in which it figures, I appeal first to a fanciful case involving a set of ever-better options, and then to a more down-to-earth case involving a vague goal, to develop a defense of the opposing view that, even in continued endorsement cases, mourning the loss of a forgone good need not be tied to the idea that the loss of the good is not fully made up for by the gain of a preferred or incomparable good of a different kind.

I. This paper concerns regret, where regretting is to be understood, roughly, as mourning the loss of a forgone good. My ultimate aim is to add a new dimension to the existing debate concerning the internal logic of regret by revealing the significance of certain sorts of cases—including, most interestingly, certain down-to-earth cases involving vague goals—in relation to the possibility of regret in continued endorsement cases. By 'continued endorsement cases,' I mean cases wherein the agent does not see her prior choice as mistaken or have any relevant intervening change in preferences.

I proceed by first delving into existing debate regarding regret in continued endorsement cases. In a stock case of the relevant sort, an agent chooses, say, one travel opportunity over another, continues to endorse his choice, and yet mourns the loss of what he is missing in having forgone the other option. Intuitively, it might

C. Andreou (✉)
Department of Philosophy, University of Utah, Salt Lake City, USA
e-mail: c.andreou@utah.edu

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*, Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_4

seem like, in such cases, an agent's regret (if it is to make sense) must be tied to the idea that the forgone good is no better than the achieved good but is also not fully made up for by the achieved good because the goods are (too) different in kind. But this view is controversial. After describing a challenge to the view, as well as the main features of the debate regarding regret in which it figures, I will develop a defense of the opposing view that, even in continued endorsement cases, mourning the loss of a forgone good need not be tied to the idea that the loss of the good is not fully made up for by the gain of a preferred or incomparable good of a different kind.

My defense picks up on a connection between regret and the intricacies of effective holistic decision-making in certain challenging choice situations in which following one's preferences at each choice point (or between each option and the next available option) can lead to trouble. I focus first on a situation involving a set of ever-better options, and then turn to a more down-to-earth case involving a vague goal. In the first, very fanciful case, an immortal agent must decide when to drink a bottle of wine that keeps improving with age (Pollock 1983). In the second, more down-to-earth case, an agent who does not want to overindulge must decide when to stop eating from a family-sized bag of chips. In each case, the details of the situation I describe are such that the following hold: First, the choice situation precludes the possibility of the agent's arriving at an option that is optimal relative to her preferences. And, second, in order to avoid realizing an outcome that is unacceptable relative to the agent's "rationally innocent" concerns, the agent must settle, somewhat arbitrarily, on an option that, though acceptable, is dispreferred in relation to another option.[1] Where such settling is required, regret on the part of the agent is appropriately grounded in the idea that, although her choice was defensible relative to her concerns, she deprived herself of a preferred alternative. Here, unlike in the cases focused on in prior debate regarding regret in continued endorsement cases, the agent's loss of the forgone good cannot be softened by the thought that the loss was necessary for the gain of a preferred or incomparable good. After responding to potential objections to my reasoning, I conclude that, given the possibility of vague goals (which often cannot be sharpened without putting weight on artificial distinctions that distort the agent's original concerns), mourning the loss of a forgone good in relation to a choice that is still endorsed need not, even in down-to-earth cases, be tied to the idea that the loss of the good is not fully made up for by the gain of a preferred or incomparable good of a different sort.

II. My inquiry begins with the following question, which figures centrally in debate regarding regret in continued endorsement cases:

Q: Can regret be appropriate in continued endorsement cases—and so even apart from any supposition that one's choice was mistaken—given a monistic theory of the good?

[1] The quoted phrase is borrowed from Tenenbaum and Raffman's discussion of vague projects in "Vague Projects and the Puzzle of the Self-Torturer" (2012).

According to monistic theories of the good, when it comes to evaluating potential actions, there is only one good-making property, and so all goods are goods of the same kind.

Q is closely related to this further question:

Q*: Can regret be appropriate in continued endorsement cases—and so even apart from any supposition that one's choice was mistaken—if it is clear in the case at issue that only one good-making property was at stake and so all the goods involved were goods of the same kind?

Though I will largely focus on Q, my response to that question, which will be affirmative, will favor an affirmative response to Q* as well, and so will be of interest even to those who find monistic theories of the good implausible but allow that there are particular cases in which only one good-making property is at stake. Relatedly, my response will be of interest to any theorist concerned with the question of whether mourning the loss of a forgone good in continued endorsement cases must be tied to the idea that the loss of the good is not fully made up for by the gain of a preferred or incomparable good of a different sort. And, indeed, it is this last question, which is of interest even if one interprets all cases of regret as cases in which there is more than one good-making property at stake, that I ultimately want to address.

III. Let me begin by allowing that there is (or at least may well be) a sense of *regret* according to which regretting something necessarily involves thinking or feeling that one should have chosen otherwise, and rationally regretting something necessarily involves recognizing one's prior choice as mistaken. This is *not* the sense of regret at issue in this paper. As indicated above, regretting is here to be understood as mourning the loss of a forgone good. And, at least for the sake of argument, I will accept the philosophically common view that, given a plurality of distinct kinds of goods, there are plenty of cases of rationally regretting that do not involve the agent seeing her prior choice as mistaken.[2] (If one finds the philosophical use of "regretting" in play here unintuitive, one can replace "regretting" with "mourning the loss of a forgone good" so as to avoid terminological dispute.) Suppose, toward illustrating the philosophically common view under consideration, that parental cheerleading and professional development are two distinct kinds of goods. Then one may rationally regret missing one's child's performance even though one correctly believes that one made the right decision in taking advantage of a competing professional opportunity. Though the goods at stake may be comparable, they are not fungible.[3]

Notice that the sort of mourning I have in mind need not be deep and life-altering. Mourning the loss of a forgone good is compatible with promptly moving on with one's life—though it is also, of course, compatible with being haunted by the loss. Mourning might even be more intellectual than felt, as when one mourns the death

[2]For some relevant discussion, see, for example, (Williams 1973), (Hurley 1989: Chap. 9), (Stocker 1990: Chap. 8), and (Dancy 1993: Chap. 7).

[3]Two goods are fungible if some amount of one figures as a "perfect substitute" for some amount of the other, in the way that, to borrow from Richard Yetter Chappell (2015: 326–7), two ten-dollar bills (typically) figure as a perfect substitute for one twenty-dollar bill.

of a stranger at a public memorial, registering the loss as regrettable, even if not personally saddening.

IV. But can one rationally mourn the loss of a forgone good in continued endorsement cases if it is granted that there is only one good-making property and so all goods are goods of the same kind? Consider (a slightly modified version of) Martha Nussbaum's buttered bagels case (1986, 115–16), which seems like a forceful example in favor of the view that the answer is "no."[4] Suppose that one has to choose between a plate containing one buttered bagel and another plate containing two buttered bagels. The bagels are all "the same variety, equally fresh, equally hot," etc. (115). You will get twice as much pleasure eating the double serving, and all that matters is pleasure. If you rationally choose the plate with two bagels (and there are no intervening relevant changes), could you rationally regret missing out on the good you could have had had you chosen the plate with one bagel? The answer, it is widely agreed, is "no." There is nothing of value you missed out on that you could have had by opting for the plate with one bagel. You got all the pleasure you would have gotten choosing that option, and then some. Of course, you might wish you had had the option of having all three bagels, but having three bagels was never an option and so it is does not figure as a forgone good.

Now consider the following more general argument against the possibility of the sort of monistic regret considered in question Q:

> Suppose an agent rationally mourns the loss of a forgone good.
>
> There are two possibilities:
>
> 1. The agent sees the forgone good as preferable to the achieved good and so no longer endorses her prior choice.
> 2. The agent does not see the forgone good as preferable to the achieved good, but does not see the loss of the forgone good as fully made up for by the gain of the achieved good.
>
> In 2, the agent's regret presupposes a pluralistic theory of the good.
>
> So, in continued endorsement cases, regret for a forgone good can be rational only given a pluralistic theory of the good.

Though this argument has some intuitive appeal, existing debate regarding the internal logic of regret raises some interesting complications, particularly in relation to the idea that, in 2, the agent's regret presupposes a pluralistic theory of the good.

Consider Thomas Hurka's influential suggestion that goods with different "intrinsic properties" can be distinct "in the way that matters for rational regret" without being goods of distinct kinds, and so "monism, [like pluralism], allows for rational regret," even in cases where the agent continues to endorse her prior choice (Hurka 1996: 566). Return to Nussbaum's buttered bagel case. According to Hurka, since the bagels in Nussbaum's example all have "the same tastes, textures, and smells, only with a different causal origin," the pleasure you forgo is not "intrinsically distinct" from the one you enjoy (566). But there can be cases in which a forgone pleasure

[4]In Nussbaum's original version, the agent's "rational principle" is to "maximize her bagel eating" (115).

is intrinsically distinct from one you enjoy, even though the same good (namely, pleasure) is at stake. Suppose, for example, that you have to choose between a plate containing one savory bagel and a plate containing two sweet bagels, and you chose the plate containing the two sweet bagels. The forgone pleasure of the savory bagel is intrinsically distinct from the pleasure of the sweet bagels, and so Hurka's position allows that, even if you get more pleasure from eating the two sweet bagels than you would from eating the one savory bagel, there is room for rational regret. In short, even while granting that regret would not be rational in Nussbaum's buttered bagel case, Hurka resists skepticism concerning monistic regret (of the sort considered in question Q) by distinguishing between goods of distinct kinds and goods with different "intrinsic properties."

Here skeptics about monistic regret are likely to pursue one of two strategies: On the one hand, they might maintain that if pleasure really is all that matters, then there is nothing of value you missed out on that you could have had by opting for the plate with one savory bagel. You missed out on a certain experience, call it the experience of savoriness, but insofar as what matters is not savoriness itself but the accompanying or supervening pleasure, the fact that the pleasure you enjoyed was not tied to savoriness is insignificant; just as it is insignificant, in Nussbaum's original case, that the pleasure you enjoyed was not tied to the forgone bagel. On the other hand, skeptics about monistic regret might maintain that, insofar as one convincingly makes a case for the view that there is something of value you missed out on that you could have had by opting for the plate with one savory bagel, one thereby convincingly makes a case for the view that it is not just pleasure that matters, but that there are different kinds of pleasure, each with its own distinct sort of value; we are thus pushed to pluralism about the good. Michael Stocker (1990: 268) highlights these two possibilities when he says that

> some may not see… differences [in gustatory pleasure] as allowing for rational conflict, since they do not see it as rational to be concerned about what sort of gustatory pleasure one gets. But even if there can be rational conflict between such pleasures, what allows for the rationality of the conflict—the difference in these pleasures as pleasures—also give an evaluative pluralism of pleasure.

V. Rather than attempting to settle this dispute, which is really more about how to understand pluralism than about how to understand regret, I will focus on developing an alternative, and alternatively motivated, defense of monistic regret (with respect to a choice that is still endorsed). While Hurka's defense is meant to undermine the view that "regret for a forgone lesser good can be rational only given a pluralistic rather than a monistic theory of the good" (1996: 555), my defense will *not* appeal to cases in which the forgone good is a *lesser* good than the achieved good; relatedly, my reasoning will not rule out the view that, insofar as regret for a forgone *lesser* good can be rational, it must be accounted for by pluralism. My aim is not to settle debate on this issue, but to complicate existing debate concerning the internal logic of regret by exploring interesting cases of an altogether different sort, namely cases in which the agent sees the forgone good as preferable to the achieved good and yet does *not* see her prior choice as mistaken and continues to endorse it. In such cases, the forgone good need not even have different intrinsic properties than the achieved good for regret to be rational; relatedly, in such cases, mourning the loss of the forgone good is tied to certain intricacies of effective holistic decision-making that cut across monism and pluralism about the good.

To begin with, I turn from buttered bagels to "EverBetter Wine." Suppose that there is an immortal agent with, as in John Pollock's example, "a bottle of EverBetter Wine which continues to get better forever. When should [she] drink it?" (Pollock 1983: 417). (Note that instead of focusing on an immortal agent with a bottle of EverBetter Wine, we could instead focus on a mortal agent faced with an infinite set of increasingly better wines; but I'll stick with the original, diachronic example.)

I will assume that the agent correctly sees never drinking the wine as unacceptable, perhaps because it is her only potential source of pleasure. I will also assume that since never drinking the wine is unacceptable, and since there is no optimal drinking point, there must be one or more points in time during which drinking the wine is rationally permissible even though drinking the wine at any such point in time is not optimal.[5] This accords with the plausible assumption that rationality cannot require the impossible and so cannot, assuming the agent has not already gone astray, prohibit every option. If one insists that every option may be rationally prohibited, read "rationally permissible" as "not rationally mistaken," where an option counts as *rationally mistaken* (given the alternatives, the choice situation, and any relevant features of the choosing agent) only if the following condition obtains: A rationally well-constituted agent (with a clear grasp of her reasons for action and no choice but to pick from among her options, even if only by default) would definitely *not* pick that option.

Now suppose that all that matters is the pleasure one will get from drinking the wine, and that the better the wine, the greater the magnitude of the pleasure one will experience. Suppose also, at least for the sake of argument, that no matter when one drinks the wine, one will experience the same kind of pleasure, with all that varies being the magnitude of the pleasure experienced. Finally, suppose that the agent opts to drink the wine at time t, which is a rationally permissible drinking point. Is there room for the agent to mourn the loss of the (greater) pleasure she would have experienced had she opted to drink the wine at t+1? I take it that the answer is clearly "yes" and that we now have a clear case of the sort needed to provide an affirmative response to our initial question Q about monistic regret. In this case, such mourning seems particularly appropriate, since one's loss, though defensible given the challenge posed by the case in relation to effective holistic decision-making, is not even softened by the gain of a preferred or incomparable good of another sort, as might have been the case were a plurality of goods involved.

It might be suggested that, although Pollock's case is theoretically interesting, since it can be used to establish the possibility of monistic regret (with respect to a

[5]For a way of understanding the optimizing conception of rationality that allows for this, see, for example, (Mintoff 1997: Sect. 4), wherein Mintoff argues that "the optimising theory does *not* imply…that if one knows there is a better alternative to some action, then one ought not to perform that action" (119).

choice that is still endorsed), it is irrelevant in relation to what is actually possible for mere mortals like ourselves. (This concern regarding the potential irrelevance of the case seems particularly pressing if, in addition to not needing to worry about challenges that go along with immortality, we also do not need to worry about ever having to choose from infinite sets of increasingly attractive options.) But consideration of Pollock's case paves the way to the following important result regarding more down-to-earth cases: If one allows for vague goals (which often cannot be sharpened without putting weight on artificial distinctions that distort the agent's original concerns), then, *even in cases where there is a finite set of options,* mourning the loss of a forgone good (in relation to a choice that is still endorsed) need not be tied to the idea that the loss of the good is not fully made up for by the gain of a preferred or incomparable good of a different sort; it can instead be tied to the intricacies of effective holistic decision-making and, relatedly, to the idea that, though one's choice was defensible relative to one's concerns, one deprived oneself of a preferred alternative.

Here is a case of the sort I have in mind:[6] You have access to a family-sized bag of chips that will not be whisked away for another hour. You value both the benefits that come with eating chips (including, say, the pleasure of biting into a savory treat) and the benefits that come with not overindulging (including, say, the satisfaction of having exhibited temperance). Notice that, whether or not these goods are of distinct kinds, for any number n, such that n is greater than 0 but less than the total number of chips in the bag, eating n chips from the bag seems like the same kind of good as eating n+1 chips from the bag; and so the choice between these two options will not be construed as a choice between goods that are significantly different in kind. Suppose, relatedly, that your goal of not overindulging is vague in the sense that there is no sharp crossover point in the process of eating chips one at a time that takes one from not overindulging to overindulging (where the fuzzily-bounded range of qualifying cases, all of which you aim to avoid, can span from mild to egregious).[7] Suppose also that, as a result (and given your appetite for chips), you find that, for each n greater than or equal to 0 but less than the total number of chips in the bag, you prefer eating a total of n+1 chips over eating a total of n chips (and this holds whether or not you are, at the moment, entertaining other stopping points).[8] This is because, by hypothesis, having just one more chip will be enjoyable and will not itself make

[6] I discuss many cases of the relevant sort in my prior work on temptation and choice over time. The case I will lay out here is a variation on my fun-size cakes case in (Andreou 2014).

[7] For interesting discussion concerning the philosophical issues surrounding the idea of vague goals, see (Tenenbaum and Raffman 2012). That discussion provides compelling support for taking cases like the one I am now considering at face value. I will provide some additional support for taking the cases at face value below. Note that, as my parenthetical remark in the text is meant to flag, I think that overindulging can be properly understood as both vague and graded. In a *Pea Soup* post, Chappell (2016) suggests that goals or ends that are described as vague are usually better understood as simply graded. But his position is controversial (as suggested by some of the responses to his post), and he acknowledges that vague goals may indeed be possible, so considering the significance of vague goals in relation to rational regret is, I think, still very much in order.

[8] In particular, knowing (for some given n and n+1) that there are other stopping points available does not eliminate or reverse your preference for eating a total of n+1 chips over eating a total of n chips, even if it complicates matters given that, as will become apparent, this preference combines

Fig. 1 Read "<" as "is preferred less than." (Variations on this figure appear in my prior work on temptation and choice over time. See, especially, (Andreou 2014).)

the difference between your overindulging and your not overindulging; accordingly, eating a total of n+1 chips rather than a total of n chips has a determinate advantage relative to your concerns and no determinate disadvantage (which easily explains your preference for eating a total of n+1 chips rather than a total of n chips). Yet you also prefer having no chips at all (though you see this as a bad option) over eating all the chips in the bag (which would qualify as a clear case of overindulging, and so as a terrible option). For the moment, I will simply assume that these preferences, which form the preference cycle in Fig. 1, are not only possible and understandable, but are also not precluded from being rationally acceptable just because they are cyclic. I will return to the latter, controversial assumption below.

Relative to the preferences under consideration, there is no optimal stopping point, though there is a fuzzily bounded range of good stopping points somewhere beyond the bad option of having no chips at all but before the terrible option of eating the whole bag (and so within the range of options for which it is true that *adjacent* options are not significantly different in kind). Suppose, without loss of generality, that the good stopping points are in the ballpark of a couple of handfuls of chips. And suppose, finally—and in accordance with the appearance that your vague goal of not overindulging is "rationally innocent" (to which I will also return below)—that, as in the case of the EverBetter Wine, rationality requires you to settle, somewhat arbitrarily, on a good stopping point, and that, recognizing this requirement, you do so, with the result that you stop after some relatively small number of chips, say k. (Notably, this assumes that you, like the agent who drinks the wine at time t in my

with your preferences over other pairs of options in a way that can leave you stumped about where to stop.

elaboration of Pollock's example, have enough self-control to stick to an intention if you think it is rational to do so.)

Here, as in the wine case, and despite various important structural differences, regretting the loss of the extra pleasure you would have gotten had you drawn the line one step further (at k+1 rather than k) makes sense given your postulated concerns and choice situation, even if you accept your prior choice as in no way mistaken and plausibly take it that you are rationally required to settle on a dispreferred alternative. By hypothesis, your choice is somewhat arbitrary, since there is no optimal stopping point—or, in the wine case, unstopping point—relative to your preferences. In both cases, losing the extra pleasure that you could have had by drawing the line one step further is avoidable and dispreferred, and this seems to make regret even more understandable than in cases where the loss was necessary for a preferred or incomparable good of a different sort. (Of course, desisting from eating chips at some point early on is necessary for the good of avoiding overindulging; but, since there is no sharp crossover point in the process of eating chips one at a time that takes one from not overindulging to overindulging, one's choice is never as neat as 'at n chips or never' (though, eventually, it will be clearly too late for success).)

Now it might be suggested that, because vague goals like the goal of not overindulging prompt cyclic preferences, such goals are not rationally acceptable. Of course, this follows only if cyclic preferences are not rationally acceptable. If the acceptability of cyclic preferences is treated as an open question, then the fact that the vague goal of not overindulging seems "rationally innocent" speaks in favor of the acceptability of cyclic preferences. Is there a compelling reason to resist the rational acceptability of cyclic preferences? Borrowing from my previous work on the topic, I will here briefly explain why I think that the answer is no.

The most prominent argument against the rational acceptability of cyclic preferences is the money-pump argument. Very roughly put, the general 'moral' of the argument is that an agent with cyclic preferences can, if she follows her preferences, end up with an option that is clearly rationally unacceptable given the options she passed up on the way. In the illustrations that account for the argument's name, the agent ends up, after a series of transactions in which she follows her preferences, with the option she originally had, minus a substantial sum of money. There is no need to get into the details of the argument (or refined variations of it), since the chip case laid out above illustrates the general 'moral' of the argument well enough. If the agent in the chip case follows her preferences (and eats each chip she has the opportunity to eat, with the recognition that, however many other chips she eats in all, eating one more chip will be enjoyable and will not take her from not overindulging to overindulging), she will certainly overindulge—an option that is, by hypothesis, rationally unacceptable because it fails to meet the agent's goal. But importantly, the chip case also highlights something the money-pump argument neglects, namely that rather than prohibiting agents from having cyclic preferences, rationality may simply require agents to keep track of how their choices 'add up' and (insofar as the agent has enough information to go on) to choose in a way that avoids rationally unacceptable options even if this involves choosing somewhat arbitrarily because there is no optimal option (relative to the agent's preferences).[9]

At this point, it might be suggested that the lack of any optimal option (even though the set of options is finite) is what makes cyclic preferences irrational. But this is to assume (rather than argue) that cyclic preferences are irrational, and the same holds for the similar suggestion that, if a set of preferences are rational, there must be an optimal option, even if it is difficult to identify (Andreou 2016). The money-pump argument avoids this question-begging move, since the option deplored as clearly rationally unacceptable in the challenging illustrations the argument incorporates is not merely dispreferred—it is also, crucially, identical to another option except that it is inferior in one respect (Andreou 2016). It is this latter feature (which is *not* true of every option in money-pump cases)—rather than the option's mere sub-optimality (a feature that is shared with the original option, which is *not* assumed to be rationally unacceptable)—that makes it clear that the agent's preferences are leading her astray, and so makes the argument genuinely challenging, rather than simply question-begging. But, in avoiding the question-begging move under consideration, the argument opens itself up to the retort raised above, namely that, even if the general 'moral' of the argument is correct, one need not conclude that cyclic preferences are rationally unacceptable. Rather than taking vague goals and the cyclic preferences they can prompt as a sign of irrationality, they can instead be taken as calling for effective holistic decision-making, a challenge that all of us are familiar with and that the functional among us routinely meet.

Given the preceding reasoning, I do not think there is compelling reason to resist the rational acceptability of cyclic preferences or the rational "innocence" of vague goals that can prompt cyclic preferences. It is, however, worth emphasizing that, putting aside the question of whether or not vague goals that prompt cyclic preferences can be rationally acceptable, there is nothing about the internal logic of regret that precludes an agent with vague goals from experiencing the sort of regret I have been describing, and from the regret being rational in the sense of being appropriate relative to the agent's concerns. This is important, since exploring cases of regret that are rational in this sense seems sufficient for illuminating the internal logic of regret; and this is the central task that I hope to have achieved in my inquiry regarding whether mourning the loss of a forgone good, in relation to a choice that is still endorsed, must be tied to the idea that the loss of the good is not fully made up for by the gain of a preferred or incomparable good of a different sort.

VI. I began with the question of whether mourning the loss of a forgone good can be appropriate in continued endorsement cases, and so even apart from any supposition that one's choice was mistaken, given a monistic theory of the good. I argued that the answer is "yes." My reasoning develops a defense of monistic regret, with respect to a choice that is still endorsed, that does not depend on the influential but controversial claim that different "intrinsic properties" can be distinct "in the way that matters for rational regret" without being goods of distinct kinds. Appealing to Pollock's EverBetter Wine example, I zeroed in on a case in which there is clearly only one

[9]For some relevant distinctions and reasoning, see Andreou (2007, 2015).

kind of good at stake and in which mourning the loss of the forgone good seems particularly appropriate because the loss, though defensible given the challenge posed by the case in relation to effective holistic decision-making, is not even softened by the gain of a preferred or incomparable good of another sort, as might have been the case were a plurality of goods involved. I then considered a more down-to-earth case that supports the further conclusion that, if one allows for vague goals, then even in cases where there is a finite set of options, mourning the loss of a forgone good, in relation to a choice that is still endorsed, need not be tied to the idea that the loss of the good is not fully made up for by the gain of a preferred or incomparable good of a different sort; it can instead be tied to the intricacies of effective holistic decision-making and, relatedly, to the idea that, though one's choice was defensible relative to one's concerns, one deprived oneself of a preferred alternative.

Acknowledgments My thanks to Richard Dietz, Dan Lassiter, Mariam Thalos, and Mike White for helpful feedback on earlier drafts of this paper. I am also grateful for the helpful anonymous comments I received and for supporting research funds from the University of Utah and from the Charles H. Monson Esteemed Scholar Award.

References

Andreou, C. (2016). Cashing out the money-pump argument. *Philosophical Studies, 173,* 1451–1455.
Andreou, C. (2015). The real puzzle of the self-torturer: Uncovering a new dimension of instrumental rationality. *Canadian Journal of Philosophy, 45,* 562–575.
Andreou, C. (2014). Temptation, resolutions, and regret. *Inquiry, 57*(3), 275–292.
Andreou, C. (2007). There are preferences and then there are preferences. In B. Montero & M. D. White (Eds.), *Economics and the Mind* (pp. 115–126). New York: Routledge.
Chappell, R. Y. (2016). Do we have vague projects? *Pea Soup: Philosophy, Ethics, Academia*. http://www.peasoup.us/2016/08/do-we-have-vague-projects/.
Chappell, R. Y. (2015). Value receptacles. *Noûs, 49*(2), 322–332.
Dancy, J. (1993). *Moral Reasons*. Oxford: Blackwell.
Hurka, T. (1996). Monism, pluralism, and rational regret. *Ethics, 106*(3), 555–575.
Hurley, S. (1989). *Natural reasons*. New York: Oxford University Press.
Mintoff, J. (1997). Slote on rational dilemmas and rational supererogation. *Erkenntnis, 46*(1), 111–126.
Nussbaum, M. C. (1986). *The fragility of goodness*. Cambridge: Cambridge University Press.
Pollock, J. L. (1983). How do you maximize expectation value? *Noûs*, *17*(3), 409–421.
Stocker, M. (1990). *Plural and conflicting values*. Oxford: Clarendon Press.
Tenenbaum, S., & Raffman, D. (2012). Vague projects and the puzzle of the self-torturer. *Ethics, 123,* 86–112.
Williams, B. (1973). Ethical consistency. *Problems of the Self*. Cambridge: Cambridge University Press.

Part II
Rationality in Vague Language Use and Cognition

The Elusive Benefits of Vagueness: Evidence from Experiments

Matthew James Green and Kees van Deemter

Abstract Much of everyday language is vague, even in situations where vagueness could have been avoided (i.e., where vagueness is used 'strategically'). Yet the benefits of vagueness for hearers and readers are proving to be elusive. We discuss a range of earlier controlled experiments with human participants, and we report on a new series of experiments that we ourselves have conducted in recent years. These experiments, which focus on vague expressions that are part of referential noun phrases, aim to separate the utility of vagueness (as defined by the existence of borderline cases) from the utility of other factors that tend to co-occur with vagueness. After presenting the evidence, we argue that it supports a view where the benefits that vague terms exert are due to other influences, and not to vagueness itself.

1 Introduction

In academic use, the word 'vagueness' has a specific meaning. Keefe and Smith, for example, state that 'vague predicates have borderline cases, have fuzzy boundaries, and are susceptible to sorites paradoxes' (Keefe and Smith 1997, p. 4), as do Egré and Klinedinst (2011). The crucial criterion is the existence of borderline cases: 'a word is precise if it describes a well-defined set of objects. By contrast, a word is vague if it is not precise' (Lipman 2009, p. 1). A typical example is the word 'tall', as applied to people for example, because there is no precise, known height which separates those who are tall from those who are not. The crucial point is that 'tall' admits borderline cases (i.e., people who may or may not count as tall), which are the hallmark of vagueness as we use the term.

M. J. Green (✉)
Department of Psychology, Bournemouth University, Bournemouth, UK
e-mail: mgreen@bournemouth.ac.uk

K. van Deemter
Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

Computing Science Department, University of Aberdeen, Scotland, UK
e-mail: k.vdeemter@abdn.ac.uk

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*, Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_5

Linguists, philosophers of language, and more recently game theorists have asked why natural languages contain so many vague expressions (Lipman 2000, 2009), which are often used even in situations where the speaker could have used an expression that is not vague (i.e., crisp); in these situations we say that vagueness is used *strategically*. By introducing borderline cases, these expressions create potential misunderstandings, thereby creating 'a worldwide several-thousand year efficiency loss' (Lipman 2009, p. 1). Lipman explains the point by means of a scenario in which a speaker describes a person to a hearer, who needs to identify that person in the arrivals hall of of an airport. In such a scenario, a precise description of the person's height (e.g., 'The person's height is 187.96 cm') would be more useful than a vague one ('The person is tall'). Lipman uses this scenario to explain why standard game theory models of communication (e.g., Crawford and Sobel 1982) predict that, under certain conditions, a crisp act of communication will always have more utility than a vague act that communicates the same state of affairs.

Lipman argues that the efficiency loss resulting from vague expressions would be unlikely to have arisen unless there were advantages as well as disadvantages associated with vague expressions. Lipman asks, essentially, what these advantages might be. Several tentative answers to Lipman's question have been offered (see, van Deemter 2009, 2010). One of the most promising answers appears to be the idea that vague expressions are easier to process, by a speaker and/or a hearer, than expressions that are not vague (i.e., crisp) (e.g., Lipman 2009; De Jaegher 2003; van Rooij 2003). For example, Lipman writes: 'For the listener, information which is too specific may require more effort to analyze' (Lipman 2009, p. 11). We shall refer to this as the *cost reduction* hypothesis.

This article brings an experimental approach to these issues, focussing on vagueness in descriptions (e.g., 'the square with few dots') and its effect on the hearer's ability to act on a given description, as measured by the time that it takes hearers to click on the referent of a description.[1] We find that, although hearer benefits from vague descriptions are straightforward to demonstrate in many cases, a closer experimental analysis militates against the conclusion that vagueness itself—as defined above, in terms of the existence of borderline cases—lies at the heart of these results. Instead, it is other factors, such as the presence of an overt numerical expression in the description, that proved to be decisive. We believe that, despite the fact that our experiments are unavoidably focussed on a specific class of vague expressions (since any experiment can only deal with a limited number of different stimuli), these findings are potentially important, because they call into question whether 'strategic' vagueness (i.e., vagueness where the speaker had a choice, because she could have produced a crisp expression instead) has any advantages at all. In other words, returning to Lipman's question, it is possible that vagueness has evolved partly as a necessary evil (e.g., because of the limits of observation and prediction) and partly as a side effect of other factors.

[1]Other metrics could have been chosen, such as hearers' ability to remember information, for example, or error rates. Although error rates play a minor role in the present paper, for reasons that will become clear, we focus on response times in particular.

2 Related Work

A number of answers to Lipman's question have been proposed, but as we have argued elsewhere, it appears that some of these do not to hold up to scrutiny (van Deemter 2009, 2010). For example,

- It has been proposed that the existence of vague words makes a language more efficient (Barwise and Perry 1983). The idea is that vagueness allows the use of one and the same word (e.g., 'big') in different situations, which require different standards. For instance, the vague word 'big' can denote a mouse as well as an elephant. The idea is plausible, but the problem is that 'big' is efficient not because it is vague (i.e., not because it allows borderline cases) but because it is context dependent. To see this, note for example that superlatives are not vague (i.e., they do not permit borderline cases), but they *are* context dependent: 'the biggest x' ascribes different sizes to the referent depending on x. It is for this reason that we can use the word 'biggest' to talk about the biggest mouse, the biggest elephant, and so on. Context-dependence, however, does not imply vagueness.
- Vague words are capable of combining a statement of quantity with an evaluative statement (Veltman 2002). The idea is that when we say 'The patient has a fever,' we do not merely assert that her temperature lies above a certain value, we also imply that the deviation is clinically significant (i.e., something is not right with the patient). The problem with this plausible idea is that evaluation does not imply vagueness. For example, the medical term 'obese' is evaluative (i.e., it is not healthy to be obese), yet its standard medical definition in terms of Body Mass Index is perfectly crisp (i.e., without borderline cases), with all and only BMI values above 30 counting as obese.

Analogous observations can be made about a number of other purported benefits of vagueness; one of the few hypotheses left standing at the moment is the *cost reduction* hypothesis, which we mentioned in the previous section. The findings on which we are reporting in the present paper will follow a familiar pattern: we will explore the cost reduction hypothesis, only to conclude that it may not stand up to scrutiny. Our approach, this time, will be experimental: we conduct controlled experiments to investigate the effect that vague expressions have on a hearer.

Charting the utility of vagueness is also the attested aim of a small number of previous experimental studies. But, as it happens, few of these studies have truly focussed on vagueness in the sense in which we do here (i.e., they do not address Lipman's challenge). Two recent studies illustrate this issue; let's discuss them briefly.

In a series of studies of behaviour modification, (Mishra et al. 2011) manipulated the presentation format of information about quantities in the domains of mental acuity, physical strength, and weight loss. In the weight loss study, participants were told that the study was designed to test the validity of a new (actually fictitious) health index, the HHI (Holistic Health Index). They were told that an ideal HHI score lies in the range of 45–55. In a longitudinal study, participants submitted their weight to a computer each week. Participants were told that two algorithms would be used to compute their HHI, and that the two might give different values initially, in which case the true score lay between the two values. In one condition, which the authors called the precise condition, the two algorithms gave the same score. In the other condition, which the authors called the vague condition, one algorithm added 3% to the score while the other algorithm subtracted 3% from the score, yielding a range of values whose midpoint was the same as the two values given in the precise condition.

One group of participants was given HHI scores in the ideal range: for this group their weight loss did not differ depending on whether they were given vague or precise HHI values. However for the other group, who were given HHI scores outside the ideal range, their weight loss was significantly greater if they were given vague HHI scores than if they were given precise HHI scores. The authors explain the improvement in the vague condition for this group as resulting from the participants' freedom to think of themselves a positioned on one end of the range—the end closest to the ideal HHI scores. This 'illusion of proximity' (Mishra et al. 2011, p. 4) to the goal is argued to have allowed participants to generate positive expectancies that lead to behaviours that improved performance. In contrast, in the precise conditions, participants did not have this freedom of interpretation, and could not distort the information to bring about the beneficial *illusion of proximity*. These results are interesting, and of obvious potential practical importance. We note, however, that information presented as an exact range of values does not conform with the standard definition of vagueness (Keefe and Smith 1997; Egré and Klinedinst 2011), since an exact range does not admit borderline cases. In the terminology of Hobbs (1985), the difference between a range and a single midpoint value is a difference of *granularity*. Furthermore, the experiments of Mishra et al. did not explore benefits in terms of processing cost, but in terms of long-term behaviour change.

Similar issues arise from the work of Peters et al. (2009). The authors carried out a series of studies where participants were required to rate hospitals based on various sources of information about quality of care. There was a between-subjects manipulation based on numeracy. The format of the information was manipulated within subjects: either numbers only were presented, or both numbers and evaluative categories were presented (e.g., *Poor*, *Fair*, *Good*, *Excellent*, with crisp visual boundary lines between the categories). Results showed that, for low-numeracy participants, the presence of evaluative categories resulted in a diminished influence of an irrelevant affective state on the ratings. For all participants, the presence of evaluative categories resulted in better decisions and in a greater use of the most important and reliable types of information, such as survival rates.

It is, however, questionable whether the 'evaluative categories' manipulation in this study can be considered a manipulation of vagueness. Certainly, terms like *Fair* admit the possibility of borderline cases. However, given that the boundaries between the categories were marked crisply, and that therefore the categories mapped crisply to numerical values, it becomes doubtful whether any borderline cases could be conceived to arise in fact. For example, *Fair* was mapped to 60–70% for the variable *percentage of heart attack patients given recommended treatment (ACE inhibitor)*. Accordingly, rather than the vagueness of categories such as *Poor*, Peters et al. emphasise the evaluative content inherent in these categories, and the affective potential of the evaluative content rather than the vagueness of the terms like *Fair*.

3 Our Approach to the Problem

Are vague expressions processed more easily by readers than crisp ones? Like Lipman, we focus on situations where numerical information is used in order to identify a referent. Reference, in other words, will be the communicative task on which we focus, partly because of the interest that this topic has recently drawn in various areas of Cognitive Science (van Deemter 2016). By looking at one specific type of vagueness, we will be able to investigate the costs and benefits of vagueness relatively thoroughly. Whether our findings generalise to other uses of vagueness is a question on which we will speculate in the final section of this chapter.

We have chosen a narrative strategy in which we address a sequence of four experiments with human readers chronologically, explaining how each experiment helped us refine our research question. In order to do justice to our findings, we need to describe these experiments in a fair amount of detail.

Let us start by explaining the task that was given to the participants in our experiments. We used a *speeded forced choice* task to compare the processing costs of different references to quantities. In this context, speed and accuracy of responses are the key dimensions on which the different references can be compared. Each stimulus in the experiments was a set of dot arrays containing various number of dots, together with a preceding instruction (in the form of a referring expression) to choose one of the arrays with respect to its cardinality. The participant was asked to respond as quickly as possible while avoiding errors. We manipulated the instructions and the arrays in several ways across the four experiments.

All the experiments shared the following properties: stimuli were created using the language GNU Octave (Eaton 2002) and the Psychophysics Toolbox extensions (Brainard 1997; Kleiner et al. 2007). The position of the dots was randomised per-trial. The order in which trials were presented was randomised per-participant. There were 256 trials, presented in 4 blocks of 64 each, between which the participant could rest. A MacBook Pro laptop computer with a 13-inch screen presented the stimuli to the participants and recorded responses. Participants were recruited using email lists at the University of Aberdeen, and paid ten pounds for participating. All participants self-reported fluency in English, and had normal, or corrected-to-normal vision. The experiment was conducted in a quiet room. Participants were asked to respond as quickly as possible while avoiding errors. There was a block of practice trials after which participants could ask questions, following which the experimenter left the room. All p values reported for linear models were calculated using the R package *lmerTest* (Kuznetsova et al. 2016).

When the distance grows between two numbers, they become more easily distinguishable: the *numerical distance effect* has been shown for comparing the cardinality of two sets of dots (van Oeffelen and Vos 1982) and for processing Arabic numerals and number words (Dehaene 1996). We manipulated the number of dots in each array such that some sets of arrays had smaller numerical distances and others had larger numerical distances. Where a number was mentioned in the instructions, it was always in the form of an Arabic numeral. When two numbers are presented with the smaller on the left, this left-side presentation facilitates responses indicating the smaller number: the *Spatial-Numerical Association of Response Codes (SNARC)* effect (Dehaene et al 1993; Gevers et al. 2006). We controlled which side the smaller number appeared on to avoid systematic influences of this effect.

There is abundant evidence (e.g., Trick and Pylyshyn 1994) that very small (i.e., *subitizable*) quantities are recognised and processed by a distinct psychological mechanism that differs from that used to process larger quantities. We performed a pilot experiment (Green and van Deemter 2011) in which we were able to confirm this finding in the experimental settings on which we are focussing in this paper. We found that, when participants were confronted with a stimulus consisting of two squares containing different numbers of dots.,[2] instructions of the form *Choose the square with* n *dots* led to consistently faster response times than instructions of the form *Choose the square with many/few dots* when $2 \leq n \leq 5$; the converse was true for $n > 5$. Given these findings, we henceforth focussed our studies on non-subitizable numbers, because it is there that vagueness is expected to have benefits.

In a second pilot experiment (Green and van Deemter 2013), we again presented two dot arrays with the instruction in the form *Choose the square with … dots*. The arrays contained larger numbers of dots than in the first pilot: one array always contained 25 dots, and the other contained either 5, 10, 15, 20, 30, 35, 40, or 45 dots. Each stimulus can therefore be seen in terms of the numerical difference between the number of dots in one array and the number of dots in the other: giving numerical distances of 5, 10, 15, and 20, with smaller numerical distances resulting in less discriminable arrays and larger distances resulting in more discriminable arrays. Our main manipulation was of the vagueness of the instruction, with two levels, *crisp* and *vague*. Assuming the dot array (5, 25), and the instruction referring to the smaller cardinality, the *crisp* instruction was *Choose the square with 5 dots* and its *vague* counterpart was *Choose the square with few dots*. We found, as expected, that responses were faster and more accurate for vague instructions than for crisp instructions. We also found an interaction between vagueness and numerical distance such that there were diminishing returns for vagueness as numerical distance increased, until, at the biggest numerical distance, there was no real difference any more between crisp and vague instructions. We interpreted this pattern as showing that cognitive load is relatively easy in both conditions when the stimuli are most discriminable, and so vagueness confers no additional advantage in those most easily discriminable stimuli.

However, the picture painted by these findings from the second pilot experiment might be misleading. First of all, there is a possibly confounding factor. Contrast an expression from the vague condition: ‘the square with few dots’ with an expression

[2]Such a stimulus is referred to hereafter as consisting of a set of *dot arrays* The number of dots in an array is referred to as its cardinality. The physical arrangement of dots in each array is irregular.

from the crisp condition: 'the square with 5 dots'. One difference is that 'few' has the potential for vagueness, whereas '5' is crisp. But another difference is that 'few' is verbal while '5' is numerical, in the sense that a number is mentioned explicitly. Since these two differences could not be separated in the experiment, the finding of a vagueness advantage is vulnerable to an alternative interpretation, namely that the crisp–vague difference was really an advantage for the verbal form of the quantifier. In our next experiment (reported below as Experiment 1) we therefore created verbal and numeric versions of each of the vague and crisp instructions so that we could compare the crisp–vague difference and the numeric–verbal difference in the same experiment.

In Experiment 1, reported below, we also addressed another issue with the second pilot experiment: participants chose one of two dot arrays—therefore the 'vague' quantifiers (few and many) uniquely identified one square. Recall our definition of vague—'a word is precise if it describes a well-defined set of objects. By contrast, a word is vague if it is not precise'. The quantifiers 'few' and 'many' might not have realised their potential for vagueness in the context of a choice between (only) two alternatives—where there is no borderline alternative that could result in few and many being 'not well-defined'. Furthermore, using definite articles in the instructions may have reinforced the participant's impression that only one choice counted as correct, and this impression could have been reinforced by the use of error feedback.

In the experiments we report below (Experiments 1, 2, and 3) we used what we learned from the pilot experiments to design situations that we believe did address the difference between crisp and vague instructions while taking account of the most important alternative explanations of this crisp–vague difference. The complete data and analysis for the experiments in this paper are available at `https://mjgreen.github.io/vagueness`.

4 Experiment 1: Separating Vagueness from Instruction Format

To find out what happens when words are used in a context where their potential for vagueness comes to the fore, Experiment 1 used three arrays (rather than two) so that the vague description had more than one possible referent, used indefinite articles to avoid the impression that only one response counted as correct, and was carried out without error feedback. An indication that the potential for vagueness was indeed realised in Experiment 1 is that the borderline response was chosen fairly often: 16% of the time.

In Experiment 1, an item was an instruction followed by a set of three dot arrays defined by a triple of numbers, representing the number of dots in the left, middle, and right arrays. We used four different triples of numbers: (6, 15, 24); (16, 25, 34); (26, 35, 44); (36, 45, 54). Each set of arrays comprised three arrays (instead of two as in Green and van Deemter 2013); the array representing the central number was always presented in the middle of the three; there were two flanking arrays where one had fewer dots than the central array and the other had more, and these flanking arrays appeared equally often on the left and right of the central array.

Table 1 gives the full set of stimuli and associated instructions. The way in which borderline responses were construed is as follows, using as an example the array (6:15:24) and instructions that identified the smaller flanking array (6). 6 was classified as the expected response. 15 was classified as the borderline response. 24 was classified as the extreme response.

- In the vague numerical condition the instruction was 'Choose a square with about 10 dots'—none of the arrays contained exactly 10 dots, but 10 is closer to 6 than it is to 15, making 6 a better response to that instruction, 15 a borderline response, and 24 an extreme response.
- In the vague verbal condition we used 'Choose a square with few dot'. We considered this to be equivalent in terms of which responses were expected (6), borderline (15) and extreme (24).
- In the crisp numerical condition we used 'Choose the square with 6 dots'. The smaller flanking array always contained exactly the specified number of dots. We considered this to be equivalent in terms of which responses were expected (6), borderline (15) and extreme (24).
- For crisp verbal, we used 'Choose the square with the fewest dots'. We considered this to be equivalent in terms of which responses were expected (6), borderline (15) and extreme (24).

On each trial, first the referring expression that constituted the instruction for that trial was displayed (e.g., 'Choose a square with about 10 dots'). Participants then pressed a key to indicate that they had read the instruction. The instruction remained on screen, and after 1000 ms, the arrays appeared. An example stimulus is given in Fig. 1. Response time was measured from the presentation of the arrays until the keypress indicating the participant's choice. The trial would timeout after 60 seconds if there was no response. In this experiment, no feedback was given. This was because, in the vague conditions, we did not regard any response as 'correct' or 'incorrect', but instead as 'expected response'; 'borderline response'; and 'extreme response', and we did not want to draw participants' attention to this distinction explicitly. The participant's choice was recorded for analysis.

Table 1 Experiment 1 instructions arranged by condition. The instructions given in the table started with 'Choose …'

Item	Quantity	Number	Crisp	Vague
06:15:24	Small	Numeric	The square with 6 dots	A square with about 10 dots
06:15:24	Small	Verbal	The square with the fewest dots	A square with few dots
06:15:24	Large	Numeric	The square with 24 dots	A square with about 20 dots
06:15:24	Large	Verbal	The square with the most dots	A square with many dots
16:25:34	Small	Numeric	The square with 16 dots	A square with about 20 dots
16:25:34	Small	Verbal	The square with the fewest dots	A square with few dots
16:25:34	Large	Numeric	The square with 34 dots	A square with about 30 dots
16:25:34	Large	Verbal	The square with the most dots	A square with many dots
26:35:44	Small	Numeric	The square with 26 dots	A square with about 30 dots
26:35:44	Small	Verbal	The square with the fewest dots	A square with few dots
26:35:44	Large	Numeric	The square with 44 dots	A square with about 40 dots
26:35:44	Large	Verbal	The square with the most dots	A square with many dots
36:45:54	Small	Numeric	The square with 36 dots	A square with about 40 dots
36:45:54	Small	Verbal	The square with the fewest dots	A square with few dots
36:45:54	Large	Numeric	The square with 54 dots	A square with about 50 dots
36:45:54	Large	Verbal	The square with the most dots	A square with many dots

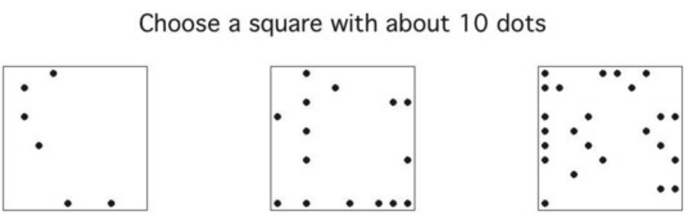


Fig. 1 Experiment 1 and 2 example stimulus

4.1 Hypotheses (Experiment 1)

We formulated the following hypotheses for Experiment 1:

Hypothesis 1 (Crisp/Vague RT) Vague instructions should result in faster responses than crisp instructions; and this pattern should hold when the model is restricted to numeric-only data and when it is restricted to verbal-only data.

Hypothesis 2 (Numeric/Verbal RT) There should be no real difference between responses to numeric instructions and verbal instructions (based on our interpretation of the experiment in Green and van Deemter (2013), where we thought that vague instructions alone were driving the advantage for instructions that were both vague, and also in verbal format).

Hypothesis 3 (Item RT) Responses should take longer as the number of dots in the display grows larger (i.e., as the levels of *Item* increase).

Hypothesis 4 (Response Type) Vague instructions should lead to more borderline responses than crisp instructions.

4.2 Results (Experiment 1)

Response Times

30 participants were recruited. Response times from all trials were trimmed at 2.5 standard deviations for each subject, leading to the loss of 236 trials, 3.1% of the data. The distribution of remaining response times was skewed with many long responses. These remaining response times were log-transformed, which reduced this skew so that their distribution more closely approximated a normal distribution. Condition means for response times are given in Fig. 2.

A linear mixed model was constructed for the (log-transformed) response times, with sum-coded vagueness, instruction format, and their interaction, and item, as fixed effects, and per-participant intercepts and slopes for sum-coded vagueness, instruction format, and their interaction as random effects.

Test of Hypothesis 1 (Crisp/Vague RT) Vague instructions actually led to significantly slower responses than crisp instructions, against Hypothesis 1 ($\beta = 0.058$, $se = 0.013$, $t = 4.55$, $p < 0.001$). When the model was restricted to numeric-only instructions, vague instructions still led to significantly slower responses than crisp instructions ($\beta = 0.093$, $se = 0.021$, $t = 4.51$, $p < 0.001$). When the model was restricted to verbal-only instructions, vague instructions tended to slow responses, but not significantly ($\beta = 0.024$, $se = 0.016$, $t = 1.46$, $p = 0.155$).

Test of Hypothesis 2 (Numeric/Verbal RT) There was actually a significant difference between numeric and verbal instructions, with numeric instructions leading to longer responses than verbal instructions, against Hypothesis 2 ($\beta = 0.265$, $se = 0.072$, $t = 5.08$, $p < 0.001$).

Test of Hypothesis 3 (Item RT) Responses took longer as the levels of *Item* increased, supporting Hypothesis 3 ($\beta = 0.120$, $se = 0.017$, $t = 7.11$, $p < 0.001$).

However, given that the response time plot in Fig. 2 shows that responses to 6:15:24 in the 'crisp numeric' instructions condition were extremely fast relative to the 'vague numeric' instructions to 6:15:24, the effects in the model of the full dataset could be driven by this difference. A clearer picture of the effects of interest might be obtained by removing the 6:15:24 level of Item from the data set, and fitting the model to this restricted data. Doing this did not affect the direction of the effects in the full dataset, but whereas the effects were significant in the full dataset, they were not significant in the restricted dataset. Full details of the analysis of the restricted dataset are available in the online materials.

Borderline Cases

A generalized linear mixed model (Jaeger 2008) was fit to the data for the distribution of responses indicating the borderline response, with sum-coded vagueness, instruction format, (and their interaction), and item as fixed effects, and the same effects as slopes over participant, as well as per-participant intercepts as random effects. The distribution of responses over the nearest match square, the borderline square, and the furthest match square are given in Fig. 3. Participants chose the borderline square on 16.6% of trials overall.

Test of Hypothesis 4 (Response Type) Participants were significantly more likely to choose the borderline option for vague instructions than for crisp instructions (21.9 versus 11.3%: $\beta = 0.62$, $se = 0.22$, $z = 2.8$, $p = 0.0059$). Participants

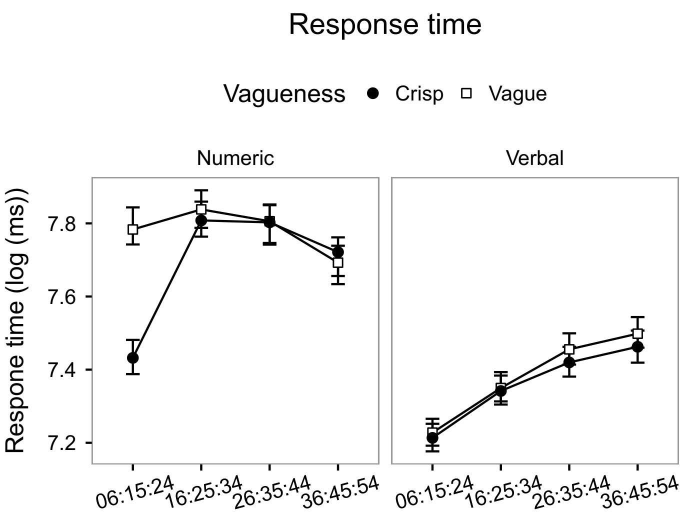
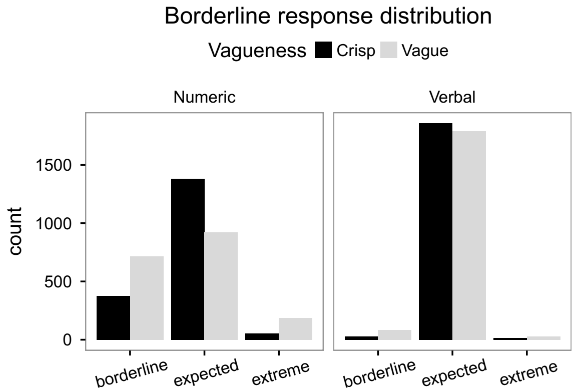


Fig. 2 Experiment 1 results: Mean response times by condition and item

Fig. 3 Experiment 1 results: Counts of borderline case responses by condition



were also significantly more likely to choose the borderline square when the instruction used the numerical format rather than the verbal format (30.1 versus 3.0%: $\beta = -3.35$, $se = 0.23$, $z = -14.6$, $p < 0.0001$).

4.3 Discussion (Experiment 1)

This experiment tested whether vague instructions would result in faster responses than crisp instructions, when borderline cases were present. Faster responses for vague instructions were found in pilot experiment B, but there were no borderline cases in that experiment.

In this experiment we found in contrast that vague instructions resulted in slower responses than crisp instructions: a difference that was significant when considering the full data (112 ms), but which was not significant after removing the smallest arrays from the analysis, which had a pattern opposite to the main trends in the rest of the data.

We also found that the effect of instruction format was significant, with numerical format slowing responses by 689 ms on average, such that the disadvantage of numerical format overwhelmed the contribution of vagueness. The verbal vague condition still yielded faster responses than the numerical crisp condition, so the pattern from pilot experiment B was reproduced, but in the light of the evidence from this experiment (Experiment 1), in the presence of borderline cases, the advantage that was ascribed to vagueness before now looks more like an advantage of verbal instruction format.

However, once again there is a possibly confounding factor. Observe that, in Experiment 1, instruction format (i.e., the difference between numeric and verbal) went hand in hand with what might be called the (human) 'selection algorithm': To see this, consider the task of selecting the dot array that contains 'few dots': to do this, it suffices to *compare* the three arrays and select the one that contains the fewest elements. To select the dot array that contains '16 dots' seems to require the participant to estimate, and then *match*, the cardinality of (at least) one dot array to 16, a process which could plausibly take longer, independently of vagueness. Therefore, our results so far permit the interpretation that what made the instructions in the verbal condition fast is not the fact that they were worded verbally, but that they allowed participants to use *comparison* rather than having to resort to *matching*.

In the next two experiments we pitted the comparison algorithm and matching algorithm selection tasks against each other while controlling vagueness and instruction format. In Experiment 2 we restricted all the instructions to numeric quantifiers while factorially manipulating vagueness and selection task. In Experiment 3 we ensured that all instructions used verbal quantifiers, while also factorially manipulating 'vagueness' and 'selection task'. This allowed us to distinguish between the predictions of the selection task account and the instruction format account.

5 Experiment 2: Focus on Instructions that Contain Numerals

The main aim of Experiment 2 was to see whether vagueness would exert beneficial effects when all conditions used numerals in the instructions, and when there were vague and crisp versions of the instructions for both comparison and matching strategies. The main changes from Experiment 1 were that the human selection task was explicitly controlled (i.e., whether the task amounted to matching or comparison), and that all conditions were constrained to mention a number. We used the same arrays as in Experiment 1 (an example stimulus is given in Fig. 1). We used a 2×2 factorial manipulation of vagueness and selection task (see Table 2). On each trial an instruction was presented: participants pressed a key to dismiss the instruction, at which time the dot arrays were presented until the participant responded, and the response time and choice were recorded. Table 2 shows the instructions for each condition. Note the difference between 'fewer than 20' and 'far fewer than 20': whereas the former cannot have borderline cases (i.e., for each number it is clear whether the number is smaller than 20 or not), the latter can.

5.1 Hypotheses (Experiment 2)

For Experiment 2, we formulated the following hypotheses:

Hypothesis 1 (Crisp/Vague RT) Vague instructions should result in faster responses than crisp instructions.

Hypothesis 2 (Comparison/Matching RT) Instructions that allow comparison should result in faster responses than instructions that necessitate matching.

Table 2 Experiment 2: Instructions arranged by condition. The instructions given in the table started with 'Choose a square with …'

Item	Quantity	Selection	Crisp	Vague
06:15:24	Small	Comparison	Fewer than 20 dots	Far fewer than 20 dots
06:15:24	Small	Matching	6 dots	About 10 dots
06:15:24	Large	Comparison	More than 10 dots	Far more than 10 dots
06:15:24	Large	Matching	24 dots	About 20 dots
16:25:34	Small	Comparison	Fewer than 30 dots	Far fewer than 30 dots
16:25:34	Small	Matching	16 dots	About 20 dots
16:25:34	Large	Comparison	More than 20 dots	Far more than 20 dots
16:25:34	Large	Matching	34 dots	About 30 dots
26:35:44	Small	Comparison	Fewer than 40 dots	Far fewer than 40 dots
26:35:44	Small	Matching	26 dots	About 30 dots
26:35:44	Large	Comparison	More than 30 dots	Far more than 30 dots
26:35:44	Large	Matching	44 dots	About 40 dots
36:45:54	Small	Comparison	Fewer than 50 dots	Far fewer than 50 dots
36:45:54	Small	Matching	36 dots	About 40 dots
36:45:54	Large	Comparison	More than 40 dots	Far more than 40 dots
36:45:54	Large	Matching	54 dots	About 50 dots

Hypothesis 3 (Interaction) The vagueness effect should differ according to whether the selection task is comparison or matching.

5.2 *Results (Experiment 2)*

38 participants were recruited. Response times from all trials were trimmed at 2.5 standard deviations for each subject, leading to the loss of 204 trials (2.8% of the trials). The distribution of remaining response times was skewed with many long responses. These remaining response times were log-transformed, which reduced this skew so that their distribution more closely approximated a normal distribution. Condition means for response times are plotted in Fig. 4.

A linear mixed model was constructed for the logged response times, with sum-coded vagueness, selection task, and their interaction, and item as fixed effects, and

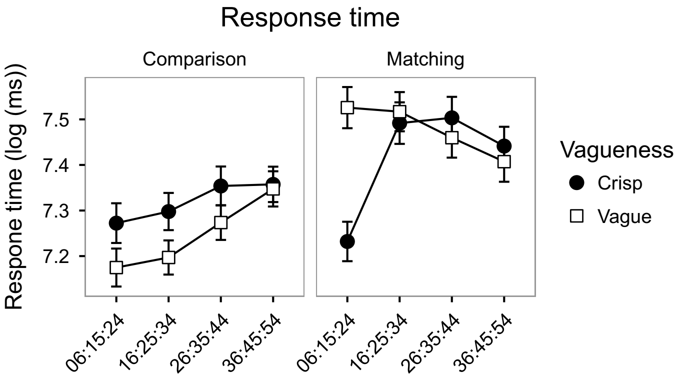


Fig. 4 Mean response times by condition for Experiment 2 where all instructions were numeric

per-participant intercepts and slopes for sum-coded vagueness, selection task, and their interaction, and item for random effects.

Test of Hypothesis 1 (Crisp/Vague RT) Vague instructions resulted in faster responses than crisp instructions on average. However this difference was not significant in the full model ($\beta = -0.0057$, $se = 0.0137$, $t = -0.42$, $p = 0.678$). Using Levy's method (Levy 2014) to test for main effects in the presence of higher-order interactions, by doing model comparison between a null model that included all interaction terms involving vagueness but leaving out a term for the main effect of vagueness, against a full model that differed only by including vagueness as a main effect, showed that the full model was no better than the reduced model ($df = 1$, $p = 0.676$), constituting more evidence that vagueness did not exert a significant main effect on response times.

Test of Hypothesis 2 (Comparison/Matching RT) Instructions involving comparison resulted in faster responses than instructions involving matching, and the difference was significant ($\beta = 0.1618$, $se = 0.0255$, $t = 6.34$, $p < 0.001$).

Test of Hypothesis 3 (Interaction) The interaction between vagueness and selection task was significant ($\beta = 0.1306$, $se = 0.0205$, $t = 6.38$, $p < 0.001$), showing that vagueness exerted different effects on response times according to whether the instructions involved comparison or matching. When separate analyses were carried out testing for the effect of vagueness in comparison-only and in matching-only conditions, vague instructions led to significantly *faster* responses than crisp instructions in the comparison-only conditions ($\beta = -0.071$, $se = 0.020$, $t = -3.52$, $p < 0.01$); and in the matching-only conditions vague instructions led to significantly *slower* responses than crisp instructions ($\beta = 0.062$, $se = 0.021$, $t = 2.91$, $p < 0.01$).

Table 3 Experiment 3 instructions for the smallest array, arranged by condition. The instructions given in the table started with 'Choose a square with …'

Item	Target	Selection	Vagueness	Instruction
06:15:24	6	Matching	Crisp	The same number of dots as the target
06:15:24	10	Matching	Vague	About the same number of dots as the target
06:15:24	20	Comparison	Crisp	Fewer dots than the target
06:15:24	20	Comparison	Vague	Far fewer dots than the target

5.3 Discussion (Experiment 2)

The cost reduction account predicted that there should be a significant main effect of vagueness such that responses would be faster for vague instructions than for crisp instructions. We found that although there was a very small effect in that direction, the effect was not statistically significant. Models that differed only in the presence of vagueness as a main effect were shown not to differ significantly in their explanatory value. However we did find that vagueness exerted effects on other variables: vagueness speeded RTs in the comparison task and slowed RTs in the matching task.

6 Experiment 3: Focus on Instructions that Do Not Contain Numerals

This experiment mirrors Experiment 2, but focuses on instructions that do not use a number. We manipulated vagueness and the selection task (comparison and matching). In order to implement the experiment without mentioning numbers in the instructions, we changed the sequence of each trial to include a 'target' (i.e., a dot array of a particular cardinality) before the array, so that we could then refer to the target's cardinality in the instruction using expressions like *the same number of dots as the target*; *fewer dots than the target*. An example of this sequence is given in Fig. 5. This presentation of a target before the main body of the trial shares some features with Izard and Dehaene (2008), Experiment 2, although in that experiment participants were told the cardinality of the target (called an *inducer* in that paper), whereas in our experiment we did not tell participants the cardinality of the prime array. An item was thus a combination of a target dot array, an instruction that did not contain a number, and a set of dot arrays taking their cardinalities from the same triples used in Experiments 1 and 2. Table 3 spells out how the instructions were constrained so as not to mention a numeral and gives examples of targets.

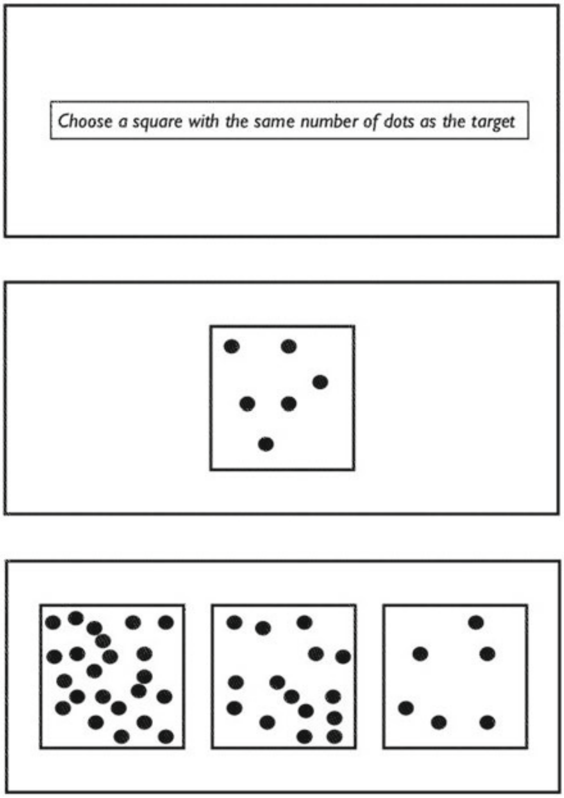


Fig. 5 Experiment 3 example stimulus, showing the sequence of presentation, with the instruction screen at the top, followed by the presentation of the 'target' array, followed by the stimulus to which participants responded

6.1 Hypotheses (Experiment 3)

For Experiment 3, we hypothesised:

Hypothesis 1 (Crisp/Vague RT) Vague instructions are easier for the reader than crisp ones.

Hypothesis 2 (Comparison/Matching RT) Comparison is easier for the reader than matching.

Hypothesis 3 (Interaction) The effect of vagueness differs depending on whether the selection task mandated by the instructions is matching or comparison.

6.2 Results (Experiment 3)

40 volunteers participated. Response times from all trials were trimmed at 2.5 standard deviations for each subject, leading to the loss of 211 trials (2.8% of the trials).

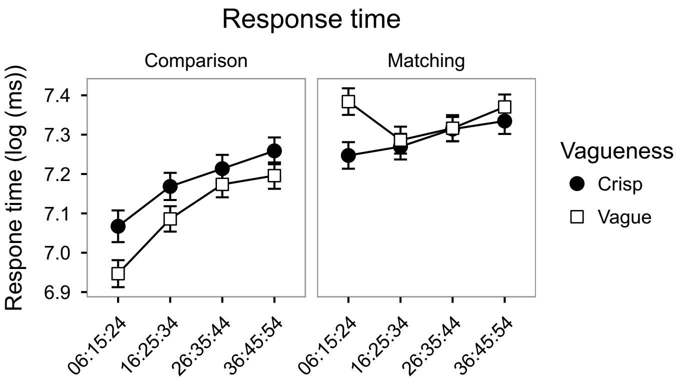


Fig. 6 Mean response times by condition for Experiment 3, where all instructions were verbal

The distribution of remaining response times was skewed with many long responses. These remaining response times were log-transformed, which reduced this skew so that their distribution more closely approximated a normal distribution. Condition means for response times are plotted in Fig. 6.

A linear mixed model was constructed for the logged response times, with sum-coded vagueness, selection task, and their interaction, and item as fixed effects, and per-participant intercepts and slopes for sum-coded vagueness, selection task, and their interaction, and item for random effects.

Test of Hypothesis 1 (Crisp/Vague RT) Vague instructions resulted in faster responses than crisp instructions on average. However this difference was not significant in the full model ($\beta = -0.015$, $se = 0.0097$, $t = -1.58$, $p = 0.119$). Using Levy's method (Levy 2014) to test for main effects in the presence of higher-order interactions, by doing model comparison between a null model that included all interaction terms involving Vagueness but leaving out a term for the main effect of vagueness, against a full model that differed only by including vagueness as a main effect, showed that the full model was no better than the reduced model ($df = 1$, $p = 0.118$), consituting more evidence that vagueness did not exert a significant main effect on response times.

Test of Hypothesis 2 (Comparison/Matching RT) Comparison instructions resulted in significantly faster responses than matching instructions ($\beta = 0.176$, $se = 0.0168$, $t = 10.51$, $p < 0.001$).

Test of Hypothesis 3 (Interaction) The interaction between vagueness and selection task was significant ($\beta = 0.123$, $se = 0.0166$, $t = 7.42$, $p < 0.001$), showing that vagueness had a different influence on response times when the instructions mandated comparison versus when the instructions mandated matching. When separate analyses were carried out testing for the effect of vagueness in comparison-only and in matching-only conditions, vague instructions led to significantly *faster* responses than crisp instructions in the comparison-only conditions ($\beta = -0.077$, $se = 0.018$, $t = -4.31$, $p < 0.001$); and in the matching-only conditions vague instructions led to significantly *slower* responses than crisp instructions ($\beta = 0.047$, $se = 0.013$, $t = 3.71$, $p < 0.001$).

6.3 Discussion (Experiment 3)

The cost reduction account predicted that there should be a significant main effect of vagueness such that responses would be faster for vague intructions than for crisp instructions. We found that although there was a very small effect in that direction, the effect was not statistically significant. Models that differed only in the presence of vagueness as a main effect were shown not to differ significantly in their explanatory value.

However, the results also showed that vagueness did exert an influence on reaction times according to whether the task was comparison or matching: vagueness was beneficial for comparison and detrimental for matching (the same as Experiment 2) even when no numbers were allowed in the instructions.

7 Discussion of Experiments 2 and 3

The main aim of these two experiments was to test whether vagueness confers any cognitive benefits over and above those that are due to differences in the selection task according to whether the instruction mandates a *comparison* selection task or a *matching* selection task, when number-use is held constant. The main effect of selection task showed that the assumption that the *comparison* task is easier than the *matching* task is well-founded. In both experiments people were reliably quicker to respond in the *comparison* task.

Vagueness, which was the phenomenon on which our investigation focussed, did not exert a significant main effect in response time. However when the comparison and selection tasks were analysed separately, there was small significant advantage for vagueness in the *comparison* tasks, but a small significant disadvantage for vagueness in the *matching* tasks (Table 4).

8 General Discussion

To summarise our findings, we asked why strategic vagueness is as frequent as it is and we decided to focus on what we see as the most promising explanation, namely that vague expressions are easy to process for speakers and hearers: the cost hypothesis, as we have called it. We decided to test this hypothesis by experimentally investigating whether vague descriptions are resolved by hearers more quickly than crisp ones. Although we were able to find some interesting (and statistically highly

Table 4 Vagueness as range reduction: A summary of Experiments 2 and 3

selection task	Vagueness	Candidates	Effect of vagueness
Comparison	Crisp	2	Vagueness advantage
	Vague	1	
Matching	Crisp	1	Vagueness disadvantage
	Vague	2	

significant) effects, it appears to us that if our sequence of experiments is assessed as a whole, it cannot be seen as confirmation of the cost hypothesis.

To explain why this is, let us summarise our findings so far: Experiment 1 showed that number avoidance in the verbal format instructions is an important factor driving the faster response times in the task, and that vagueness does not have any additional benefit in either the verbal format instructions or the numerical format instructions. However, Experiment 1 did not distinguish the benefits of number avoidance from the benefits of the comparison selection task. In Experiments 2 and 3 we manipulated vagueness and the selection task, separately at each level of numerical format. Across the two experiments, we found that the comparison-task instructions attracted faster response times than the matching-task instructions. Within the two experiments we found that vagueness exerts benefits when the selection task is *comparison*, but not when the task is *matching*.

What is one entitled to conclude? Given that we were able to identify a class of situations in which vague expressions led to faster response times than crisp ones, would it be valid to conclude that we have finally discovered an advantage for vagueness that cannot be ascribed to some other factor? We believe the answer to this question is negative.

To see why, consider Figs. 4 and 6. Both figures depict four conditions, depending on whether the expression was crisp or vague, and depending on whether the referent could be identified using a comparison strategy or not. Two of the resulting four conditions result in an expression that can denote either of two referents; the other two conditions result in an expression that can only denote one referent, with the other possible referent being a marginal candidate at best.

To see why vagueness has opposite effects depending on whether it is used in matching or comparison situations, consider the stimulus with (6, 15, 24) dots. Now compare 'Choose a square with 6 dots' with its vague counterpart 'Choose a square with about 10 dots': by adding the word 'about,' we broaden the range of squares to which the expression might be referring. On the other hand, compare 'Choose a square with fewer than 20 dots' with 'Choose a square with far fewer than 20 dots': by adding the word 'far,' we did not broaden the range of squares denotable by the expression: we narrow it down, because only some of the squares that have fewer dots may have *far* fewer dots.

The benefits of vagueness in the *comparison* task in Experiments 3 and 4 could thus be explained as differences in the number of valid targets for the expression.

This leads us to speculate that the benefit for vagueness here could be due to the vague expression foregrounding a particular valid target, while the crisp expression carries with it the additional task of distinguishing between two alternative valid targets, something we propose to call a 'range-reduction' benefit.

The observation that conditions with 1 candidate lead to shorter response times than conditions with 2 candidates is consistent with the range reduction hypothesis, but not with the idea that vagueness is beneficial. It appears, in other words, that shorter response times will only result from a vague expression if this expression leads to range reduction. Once again, it is not vagueness itself that has advantages but a phenomenon (namely range reduction) that is an automatic concomitant of vagueness in some types of situations.

Looking at the entire series of experiments, our findings suggest that the observed benefits of vague expressions may be due to factors other than vagueness: factors like avoiding numbers, permitting comparison tasks, and range reduction. The picture that is starting to emerge is subtle: on the one hand, in the situations that we have been studying, vagueness is not intrinsically beneficial. On the other hand, vague expressions frequently possess other features that *are* beneficial, and these are what give us the incorrect impression that vagueness itself is beneficial. Vagueness may thus have acquired a reputation that it does not deserve. The answer to Lipman's question, of why vagueness permeates human language (see our Introduction), may lie in a different direction after all, possibly relating to benefits for the speaker rather than the hearer.

A comparison may clarify the logic of the situation. In recent years a number of studies, focussing on red wine, have suggested that alcohol, consumed in low doses, may have health benefits. An alternative explanation, however, asserts that it is not the alcohol in the wine that is beneficial, but antioxidants from grapes. If this alternative explanation is correct, then alcohol may not be healthy after all.
Our findings suggest a re-think of the questions on which much research on the utility of vagueness rests. Years of research on the logic of vagueness—giving rise to such techniques as Partial Logic (Fine 1975, e.g.), Probabilistic Logic (Edgington 1997), and Fuzzy Logic (Zadeh 1965)—have primed the research community to expect that some special utility of vagueness is an important part of the answer; but our findings call this expectation into question.

Although our own studies in this article have focussed on vagueness in descriptive noun phrases only, it seems plausible that vagueness plays a similar role in other linguistic constructs. For example, consider reports on air temperature. Given a numerical temperature measurement or prediction, we might word it as

(a) *27.2 degrees Celsius*, or
(b) *approximately 27 degrees*, or
(c) *above 25 degrees*, or
(d) *warm*,

among other candidate expressions. Which of these descriptions is most effective, for example as part of a weather report? If the linguistic literature is to be believed, then options (a) and (c) convey crisp information, whereas (b) and (d) are vague (i.e., they permit borderline cases). In the situations studied in our own experiments and the ones discussed in Sect. 1, we found no evidence that vagueness is beneficial for hearers. Rather than asking whether a candidate expression is vague, other questions might shed more light on the choice, similar to the ones identified in our studies. These questions might focus on the amount of information that a given expression conveys (i.e., on granularity), on the avoidance of numbers, and on the use of evaluative terms. Let's see how this might pan out for the above examples from the weather domain.

First, the experiments by Mishra et al. suggest that it is important how much information is conveyed by an expression, and their findings are echoed by our own thoughts about range reduction (following Experiments 2 and 3). In the case of (a)–(d) above, it appears that (a) conveys the most detailed information (designating the smallest segment of the temperature scale), followed by (b), then (d), then (c) (e.g., 40 degrees is above 25, but at 40 Celsius the word 'warm' is likely to give way to 'hot' or 'scorching'):

$$a < b < d < c$$

If these hunches are correct, then it seems to us that it is relatively unimportant whether a given expression is vague or crisp. Other factors seem more important; moreover, it may depend on the task and the audience which of a–d is preferred. For example, an *expert* may prefer to read expression (a), because it gives her the most detailed information on which to base her decisions. On the other hand, expression (d) ('warm') is shorter than the other three and avoids the use of numbers; our experiments suggest that this may make (d) more rapidly understood than its competitors; earlier experiments point in the same direction, given the evaluative nature of 'warm' (recall section 2), which is especially important if the hearer is unfamiliar with the metric used. These considerations suggest that *non-experts* might prefer expression (d).

One way to see why vagueness (as defined in our Introduction) may not benefit human communication is the following thought experiment. Suppose a group of speakers understand the word 'warm' as vague, agreeing that temperatures above 26 count as warm, and temperatures below 24 do not count as warm, but considering temperatures between 24 and 26 as borderline cases. Now one day these speakers agree to sharpen up their definition, deciding that, henceforth, 'warm' means '$> 25°$' (as in (c) above): this decision resolves the borderline cases, while everything else remains the same. It seems unlikely that this change in language use, from a vague meaning to a crisp one (i.e., one that has no borderline cases anymore), would lower the utility of the word. Our experimental findings, and the conclusions that we draw from them, are consistent with this idea.

References

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge (Mass.) and London: MIT Press.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431–1451.

De Jaegher, K. (2003). A game-theoretic rationale for vagueness. *Linguistics and Philosophy*, *26*, 637–659.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, *8*(1), 47–68.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371.

Eaton, J. W. (2002). *GNU octave manual*. Network Theory Limited.

Edgington, D. (1997). Vagueness by degrees. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader*. Cambridge: MIT Press.

Egré, P., & Klinedinst, N. (2011). Introduction: Vagueness and language use. In P. Egré & N. Klinedinst (Eds.), *Vagueness and Language Use*. Basingstoke: Palgrave Macmillan.

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, *30*(3), 265–300.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica*, *122*(3), 221–233.

Green, M. J., & van Deemter, K. (2011). Vagueness as cost reduction: An empirical test. In *Proceedings of 'Production of Referring Expressions' Workshop at 33rd Annual Meeting of the Cognitive Science Society*. Boston, MA.

Green, M. J., & van Deemter, K. (2013). The utility of vagueness: Does it lie elsewhere? In *Production of Referring Expressions: Bridging the Gap Between Cognitive and Computational Approaches to Reference*. 31 July 2013, Berlin: Germany.

Hobbs, J. R. (1985). Granularity. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence* (pp. 432–435). Massachusetts: Morgan Kaufmann.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247.

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Keefe, R., & Smith, P. (Eds.). (1997). *Vagueness: A reader*. Cambridge: MIT Press.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception, 36*, 1–16.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). http://CRAN.R-project.org/package=lmerTest, version 2.0-32.

Levy, R. (2014). Using R formulae to test for main effects in the presence of higher-order interactions. arXiv:1405.2094.

Lipman, B. L. (2000). Comments section. In A. Rubinstein (Ed.), *Economics and language: Five essays*. Cambridge: Cambridge University Press.

Lipman, B. L. (2009). Why is language vague? Retrieved 12 April 2011 from http://people.bu.edu/blipman/Papers/vague5.pdf.

Mishra, H., Mishra, A., & Shiv, B. (2011). In praise of vagueness: Malleability of vague information as a performance-booster. *Psychological Science*, *22*(6), 733–738.

Peters, E., Dieckmann, N., Västfjäll, D., Mertz, C., Slovic, P., & Hibbard, J. (2009). Bringing meaning to numbers: The impact of evaluative categories on decisions. *Journal of Experimental Psychology: Applied*, *15*(3), 213.

Trick, L., & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*, 80–102.

van Deemter, K. (2009). Utility and language generation: The case of vagueness. *Journal of Philosophical Logic*, *38*(6), 607–632.

van Deemter, K. (2010). *Not exactly: In praise of vagueness*. Oxford: Oxford University Press.

van Deemter, K. (2016). *Computational models of referring: A study in cognitive science*. Cambridge: MIT Press.

van Oeffelen, M., & Vos, P. (1982). A probabilistic model for the discrimination of visual number. *Perception and psychophysics*, *32*(2), 163–170.

van Rooij, R. (2003). Being polite is a handicap: Towards a game theoretic analysis of polite linguistic behavior. In M. Tenneholz (Ed.), *TARK 9: Theoretical Aspects of Rationality and Knowledge*, Bloomington.

Veltman, F. (2002). *The difference between vague and not precise (Het verschil tussen vaag en niet precies)*. Vossiuspers: University of Amsterdam.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

mgreen@bournemouth.ac.uk

Towards an Ecology of Vagueness

José Pedro Correia and Michael Franke

Abstract A vexing puzzle about vagueness, rationality, and evolution runs, in crude abbreviation, as follows: vague language use is demonstrably suboptimal if the goal is efficient, precise and cooperative information transmission; hence rational deliberation or evolutionary selection should, under this assumed goal, eradicate vagueness from language use. Since vagueness is pervasive and entrenched in all human languages, something has to give. In this paper, we focus on this problem in the context of signaling games. We provide an overview of a number of proposed ways in which vagueness may come into the picture in formal models of rational and evolutionary signaling. Most argue that vague signal use is simply the best we can get, given certain factors. Despite the plausibility of the proposals, we argue that a deeper understanding of the benefits of vagueness needs a more ecological perspective, namely one that goes beyond the local optimization of signaling strategies in a homogeneous population. As an example of one possible way to expand upon our current models, we propose two variants of a novel multi-population dynamic of imprecise imitation where, under certain conditions, populations with vague language use dominate over populations with precise language use.

1 Vagueness and Rationality

The classic philosophical problem of vagueness is most starkly embodied by the sorites paradox. The original formulation is attributed to Eubulides, an ancient Megarian philosopher (Sorensen 2009), and uses the example of a heap of sand: if no removal of one grain of sand can make a heap into a non-heap, one can repeatedly

J. P. Correia—Holder of a Ph.D. studentship by the *Fundação para a Ciência e a Tecnologia* (SFRH/BD/100437/2014).

J. P. Correia (✉)
Institute of Philosophy, University of Porto, Porto, Portugal
e-mail: zepedro.correia@gmail.com

M. Franke
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
e-mail: mchfranke@gmail.com

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*,
Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_6

remove all but one grain of sand from something that is clearly a heap and be forced to acknowledge that the remaining single grain of sand is still a heap; otherwise, it seems, one would have to accept that there is a determinate number of grains that forms a heap, and anything under it is not a heap. Neither choice is, however, intuitively satisfying. The paradox is interesting because it can be made general and re-applied to many other words besides 'heap'. Predicates for which one can find a suitable instance of the general formulation of the sorites paradox are called *vague*. Paradigmatic examples besides 'heap' include 'tall', 'red', 'bald', 'tadpole', and 'child' (Keefe and Smith 1997). How widespread is the problem? It is easy to find more examples of predicates based on more finely grained properties—as 'tall' is intuitively based on height—for which constructing a sorites paradox would be easy. Mereological nihilists argue that instances of the sorites paradox can be designed for any material object that can be decomposed into small enough parts. If one subscribes to the scientific picture of matter as composed of molecules and atoms, this applies to tables and chairs, cats and mats, and any other ordinary thing (Unger 1979). Bertrand Russell famously argued (Russell 1923) that all words, including "the words of pure logic," are vague when used by human beings.

If one thinks of language as governed by logical rules, and of rationality as including the ability to follow those rules, the sorites paradox seems *prima facie* to demonstrate that vagueness and rationality are incompatible. But one can think of language in different terms. One possibility is to think of signs as tools agents use to coordinate actions. An example of a way to formally study language along those lines is the game-theoretic framework of signaling games (Lewis 1969/2011). Within such an approach, rationality can be seen as the ability to choose the use of signs that is optimal to achieve some form of coordination. The existence of vagueness in such models would not be at odds with rationality as long as vague languages turn out to be optimal for the purposes at hand. However, as Lipman (2009) argues in detail, this is typically not the case. The problem can be put very succinctly as follows. In standard game-theoretic models of communication, vague signal use yields a lower expected utility than crisp use. Therefore, given that the dynamics (be it natural selection, cultural evolution, or rational choice) maximize utility, vagueness should be weeded out by these forces, giving rise to only precise languages. Thus it would be irrational to stick to vague language use when one could (theoretically) switch to a better system. But vagueness is pervasive in natural language, and there is no reason to believe it is going away. The problem seems to be a theoretically serious one, but it does not obviously undermine our everyday linguistic practices. Most of the time we seem to communicate just fine. Therefore, the issue must lie with the conceptions we have of the forces or mechanisms that underlie those practices. Lipman concludes that "we cannot explain the prevalence of vague terms in natural language without a model of bounded rationality which is significantly different from anything in the existing literature" (Lipman 2009, p. 1).

Here we survey proposals for addressing vagueness in the context of signaling games, but before we start we want to establish some vocabulary to better frame the upcoming discussion. Rationality is an elusive notion that is frequently debated in areas like philosophy, psychology, and economics. Discussions around it usually touch on different aspects of the concept without always clearly demarcating them; this is what we want to do before going further. The aforementioned logical picture of language and meaning is focused on dichotomies like true and false, meaningful and meaningless, correct and incorrect. A sentence can be true or false only if it is meaningful, and it is meaningful if it is constructed according to correct rules. Rationality is intimately connected with the ability to follow certain procedures, not only of sentence production, but ultimately of sentence combination and reasoning (think of what is required for making a logically valid deductive inference). This is an example of what we will call a *procedural* account of rationality, one which focuses more on the means rather than the ends. One can also do the opposite and focus on the consequences instead. *Instrumental* rationality, as it is typically called, characterizes an agent's choice as rational if it maximizes the possibility of achieving a desired goal, regardless of the means. The notion is linked to Hume (1738/2005), epitomized in the following assertion: "Reason is, and ought to only be the slave of the passions." Instrumental rationality is close to a notion of rational choice that is used in economics and game theory[1]: agents are rational if and only if they make decisions that maximize their expected utility. A more in-depth discussion of the opposition in the context of theoretical economics can be found, for example, in the work of Simon (e.g. 1986).[2]

When developing models where rationality is relevant, be it constructing a logical system or setting up a formal game, the assumed epistemic relation between agents and their environment can come to bear on considerations of rationality. The verdict over how rational certain choices are can vary depending on how accurate and complete an agent's knowledge is of its environment, the goals to be achieved, the choices or rules available, the relation between those choices and the objectives, and so forth. In this respect, we will call an agent *omniscient* if it is in possession of the same information as the modeler, whereas of an agent with less than that we will say that it only has *limited* awareness of the relevant aspects of the model. Models working within the logical picture typically do not make a distinction between modeler and agent, and thus lack room to express these epistemic gaps. By abstracting away from language users, these models also typically do not represent potential interactions between them, let alone allow for repeated interaction and language change. In other types of models, however, a further aspect of rational choice needs to be considered, namely the ability of agents to make accurate predictions about how other agents behave. We can say that an agent is more or less *strategic* depending on the extent to which she is able to anticipate the actions and beliefs of other agents, and to predict medium/long term gains from repeated play. Lack of perfect knowledge of the situation or lack of ability to choose strategically can be caused by many possible factors. These include, among other things, limitations in handling information (receiving, storing, retrieving, transmitting) and limited computational resources

[1] In fact, it has been argued (Vanderschraaf 1998) that Hume's whole account of convention is very closely in line with modern game theory.

[2] Simon uses the term substantive instead of instrumental rationality, but the characterization is basically the same (Simon 1986, pp. 210–212).

to solve complex problems. We can talk about *bounded* rationality to characterize the choices of such agents. While the above definition of instrumental rationality is usually understood as requiring a single choice to maximize the expected utility in a single concrete decision situation, one might also be interested in more general choice mechanisms (e.g. Zollman and Smead 2010; Hagen et al. 2012; Fawcett et al. 2013; Galeazzi and Franke 2017). A choice mechanism is a general way of behaving for an agent involved in a variety of decision situations. When considering only a single situation, rationality can only be *local*. If we take into account the possibility of the agent's choice or choice mechanism hinging on multiple situations, we can also talk about *global* rationality. This can be important because, hypothetically, there could be choice mechanisms that are sub-optimal at a local level (for each situation) but are actually perfectly rational at a global level.

Note that we consider that, in theory, most combinations of these aspects are possible. Although we have been pinning procedural rationality to the logical picture of language, this is only with the most traditional logical systems in mind. We are not denying that advances in dynamic, epistemic, fuzzy, paraconsistent, and other types of logic could potentially enable one to capture procedural rationality with different characteristics. Game-theoretical models of language can, on the other hand, also combine local instrumental rationality with omniscient highly strategic agents. Our objective here is not to survey all the possibilities. We want to focus on signaling games as a framework for the study of language use and meaning. The vocabulary just introduced will, we hope, help inform the discussion that follows. We proceed by introducing the framework of signaling games in Sect. 2. In Sect. 3 we look into explanations of vagueness in a particular kind of signaling game. Section 4 tries to generalize the considerations of these proposals to argue for an approach to vagueness anchored in a more global notion of rationality. We propose and analyse the results of a novel multi-population model of imprecise imitation in Sect. 5, and summarize our conclusions in Sect. 6.

2 Signaling Games

Signaling games were first introduced as models of communication by Lewis (1969/2011). In order to support the idea that linguistic conventions can arise without any prior conventional activity, Lewis considers situations where agents' choices involve sending and receiving signals or messages.[3] One can think of two players with different roles. The first player, the sender, has knowledge about which of a number of possible states of affairs obtains and, depending on this information, chooses a signal to send. The second player, the receiver, has knowledge about which signal the sender chose and, based on this information, chooses one of several possible responses. A preference relation exists between responses and states of affairs, and a payoff is attributed to each player based on the choices of both. Note that Lewis

[3]These terms will be used interchangeably.

assumes that no player has any preference regarding the particular signal that is used for a given state, provided that it enables advantageous coordination with responses. Formally, in order to describe the setup all we need is to specify a set of possible states of affairs T, a probability measure P such that $P(t)$ is the probability or frequency with which $t \in T$ occurs, a set of available signals or messages M, a set of responses or actions A, and a pair of utility functions $U_{S,R} : T \times A \to \mathbb{R}$, one for the sender and one for the receiver, each of which yields a payoff value for each possible pairing of state and action. These so-called signaling problems can be seen as particular cases of coordination problems if we consider the players' choices to be of contingency plans or strategies. A sender strategy is a specification of a choice of message for each possible state of affairs. It thus describes the sender's behavior conditional on the state of affairs that obtains. A receiver strategy analogously specifies a choice of action for each possible message. Formally, what the sender chooses is a function $\sigma : T \to M$ and the receiver a function $\rho : M \to A$. The expected utility EU of a strategy can be calculated by using the utility function and aggregating payoffs for all pairings of states of affairs and actions, weighted by the probability of each state. Concretely, the expected utility of σ given ρ is $\mathrm{EU}_S(\sigma \mid \rho) = \sum_{t \in T} P(t)\, U_S(t, \rho(\sigma(t)))$, and the expected utility of ρ given σ is $\mathrm{EU}_R(\rho \mid \sigma) = \sum_{t \in T} P(t)\, U_R(t, \rho(\sigma(t)))$. As an example, consider a game with $T = \{t_1, t_2\}$, $M = \{m_1, m_2\}$, $A = \{a_1, a_2\}$, $P(t_1) = P(t_2) = 0.5$ and the following utility matrix:

	a_1	a_2
t_1	1, 1	0, 0
t_2	0, 0	1, 1

Consider sender strategy $\sigma = \{t_1 \mapsto m_2, t_2 \mapsto m_1\}$ and receiver strategy $\rho = \{m_1 \mapsto a_2, m_2 \mapsto a_1\}$. These would have an expected utility of 1 for both sender and receiver, since when t_1 obtains with probability 0.5 the sender will use m_2, and to this message the receiver will respond with a_1, which achieves a payoff of 1, and when t_2 obtains with probability 0.5 the sender will use m_1, and to this message the receiver will respond with a_2, which also achieves a payoff of 1. They also represent one of the two stable conventions in this game, the other being the pair of strategies $\sigma = \{t_1 \mapsto m_1, t_2 \mapsto m_2\}$ and $\rho = \{m_1 \mapsto a_1, m_2 \mapsto a_2\}$. Conventions of this kind in a signaling problem are what Lewis calls *signaling systems*. An example of complete miscoordination would be $\sigma = \{t_1 \mapsto m_1, t_2 \mapsto m_2\}$ and $\rho = \{m_1 \mapsto a_2, m_2 \mapsto a_1\}$. Partial coordination is achieved, for example, by $\sigma = \{t_1 \mapsto m_1, t_2 \mapsto m_1\}$ and $\rho = \{m_1 \mapsto a_1, m_2 \mapsto a_2\}$.

Lewis' account of the stability of conventions rests on what could be considered strong demands. Namely, there needs to be a state of affairs that indicates to everyone involved that a certain regularity will hold, as well as "mutual ascription of some common inductive standards and background information, strategic rationality, mutual ascription of strategic rationality, and so on" (Lewis 1969/2011, pp. 56–57). Agents are thus envisioned as omniscient and highly strategic. These requirements can seem excessive, and even more so if we consider how simple signaling systems are when compared to human languages. The models were introduced to help explain how language could have gotten off the ground as a conventional system without any sort of prior agreement. However, if one considers the circumstances of the origins of human language, it seems implausible that the agents that started making use of primordial signaling systems which (hypothetically) evolved into languages possessed such advanced rationality. Furthermore, communication through simple message exchange is something that almost all animals do: monkeys use calls, birds use singing, bees use dances, ants use pheromone trails, and so on. A plausible account of the origin of language should first explain how signaling systems could get started, without requiring high standards of rationality from the agents involved.

In order to address this problem, Skyrms (1996) proposes studying signaling problems in evolutionary terms. Rather than imagining, as Lewis does, rational agents making conscious decisions in possession of knowledge of the game and expectations of the behavior of other agents, one can imagine a simpler scenario inspired by biological evolution: there is a population of agents with biologically hardwired behaviors for engaging in interactions characteristic of a signaling problem; utility does not represent preference, but rather fitness for survival and reproduction; the make-up of the population evolves based on the relative fitness of the strategies represented in the population. Such a setup attempts to capture the main features of natural selection: in a diverse population, agents with more successful strategies thrive, while agents with less fit strategies die off. Although the inspiration for this scenario is biological evolution, similar things could be said about how ideas spread in a population of agents who can adopt or abandon them depending on how successful they prove to be (e.g. Benz et al. 2006; Pagel 2009; Thompson et al. 2016), i.e. we can interpret these notions in terms of cultural evolution (Dawkins 1978; Boyd and Richerson 1985). The principles can be captured in formal models that abstract away from details of single interactions and behavior of individual agents, for example in the replicator dynamics (Taylor and Jonker 1978). The only things relevant to this equation are the relative proportions of strategies in a given population and the utility function. Using it, one can compute which strategies evolve under which conditions.

Skyrms' evolutionary game theory approach to signaling games not only gives more plausible grounds to support Lewis' discussion of convention, it also accomplishes an important conceptual change: it moves most of the theory and mathematical formalism to the descriptive side of the investigation. Utility represents how the modeler views the signaling problem and understands the relative advantages or disadvantages of different possible strategy combinations. Dynamics describe how strategies can evolve when driven by mechanisms of utility maximization. The shift in perspective allows interpretations that accommodate limited non-strategic agents. While the general framework manages to abstract quite some details away from the formalization, it nevertheless leaves room for them, especially when it comes to the dynamics. We have already mentioned the replicator equation that can be seen as representing biological or cultural evolution, but one can also use dynamics inspired by learning mechanisms (e.g. Roth and Erev 1995), or even ones assuming a high degree of knowledge of the game and other players (e.g. Gilboa and Matsui 1991; Mühlenbernd 2011; Spike et al. 2017). This range of options goes hand in hand with a range of pictures of rationality, from nothing more than survival of the fittest in a biologically-inspired setting, to a certain degree of instrumental but limited and non-strategic rationality in the case of learning dynamics, to higher levels of rationality and even recursive strategic reasoning about the co-players' beliefs and choices. Each of these can be utilized depending on the problem that one is interested in characterizing. Thus, although Skyrms shows that high requirements of rationality are not necessary for signaling conventions to evolve, the framework does leave room for the study of linguistic interactions between highly strategic agents.

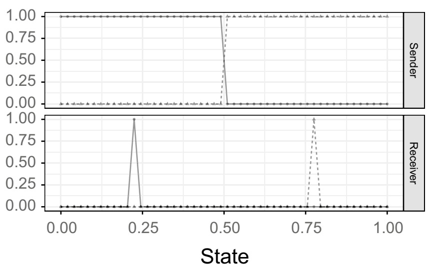
The characterization of signaling problems in terms of evolutionary game theory allows us to explain why certain equilibria come to be and how. A core notion in this context is that of an *evolutionary stable state* (Maynard Smith 1982): an equilibrium situation that a population tends to under standard evolutionary pressures, and to which it returns if slightly disturbed. With these tools, one can better understand why signaling systems are stable even without any strong assumptions of rationality. One can also map out which initial conditions drive the system towards which equilibria and which do not. In a simple case like the example discussed above, an evolutionary process of the kind described always drives the population into a state where one signaling system takes over completely. More complex signaling problems may have different evolutionary outcomes, sometimes unexpected ones. Skyrms (2010) gives an overview of different topics studied using signaling games, including expansions of the framework itself (for example, considering other dynamics beyond the replicator equation), exploration of other factors that impact the evolution of signaling (for example, how agents are interconnected), or variations on the signaling problem and its basic assumptions (for example, loosening the alignment of interests in order to provide accounts of deceptive signal use). Other uses of signaling games include discussions of categorization (e.g. Jäger and Rooij 2007), compositionality (e.g. Barrett 2009), incommensurability (e.g. Barrett 2010), to name a few. More recent overviews are given by Huttegger (2014), Huttegger et al. (2014), and Franke and Wagner (2014). In the following section, we discuss how a particular type of signaling game has been used to address the problem of vagueness.

3 Vagueness in Sim-Max Games

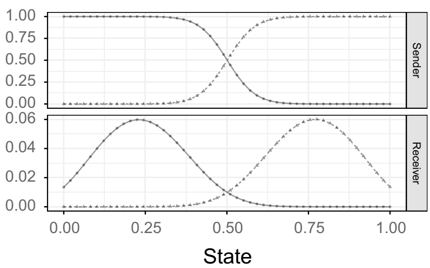
The sorites paradox requires us to assume a relation between the vague terms and a more precise underlying dimension (height for tallness, number of hairs for baldness, number of grains of sand for "heapness", and so on). Not only does this property need to be much more fine-grained than the vague term, but it also needs to have some structure: there is at least an order between the elements in it (thinking of height in centimeters, $180 > 179 > \cdots > 120$), and usually even a degree of how far apart these elements are from each other. In terms of signaling games, one can model this using a state space constituted of values of the underlying dimension, and a message space constituted by the terms in question. Because of the difference in granularity, we will typically be interested in cases where the state space is much larger than the message space. We can model the structure of the state space by defining a distance or similarity function between every value, effectively making it a metric space. Another important ingredient of the sorites paradox is the acknowledgment of a certain degree of tolerance with respect to whether a certain term applies or not. This tolerance decreases with distance in state space: assuming a 180 cm person is tall, one would easily tolerate the use of the term for a person measuring 179 cm, less so for someone who is 170 cm, and much less so for 160 cm. This can be modeled using a utility function that is continuous rather than discrete and that monotonously decreases with distance, i.e. success is not a matter of black and white, right or wrong, but a matter of degree, of how close the receiver got to the optimal response to the sender's perceived state.

The simplest type of game to study in this scenario is one where the state space and the action space are the same. We can imagine this as a game of guessing states of affairs: the sender has knowledge of a particular state, sends a message to the receiver, who in turn has to guess it; their payoff, as discussed above, is proportional to how close the guess got to the original state. These games, called similarity-maximization or sim-max games for short, were first introduced by Gerhard Jäger and Robert van Rooij (Jäger and Rooij 2007; Jäger 2007) and further studied by Jäger et al. (2011). What these authors find about this setup is that the evolutionary stable states are what they call Voronoi languages. Roughly, these are situations where the sender uses messages in a way that can be seen as partitioning the state space into convex regions, and the receiver responds with the central element of those regions.

In an abstract setup, using 50 states uniformly distributed over the unit interval and two possible messages, such an optimal language looks like what we see in Fig. 1a: at a specific point, the probability that the sender uses one message decays sharply from 1 to 0, and increases sharply from 0 to 1 for the other message; the response of the receiver for each message is a degenerate distribution over the state space which assigns all probability mass to a single state. These strategies give mutually optimal behavior for a case where the prior probability is the same for each state and utility is a linear or quadratic function of the similarity between the actual state and the receiver's guess. In general, at which point the sender switches the use of messages and which guesses of the receiver are optimal critically hinges on the priors



(a) Crisp



(b) Vague

Fig. 1 Example strategies for a state space with 50 states. Each line corresponds to a message. For the sender, it plots for each state the probability that the message is used. For the receiver, it plots for each state the probability that the response is that state, given the message over states and the utility function. Still, confirming Lipman's argument, there is no vagueness in such optimal languages.

The sender/receiver strategy pairs that we are looking for look more like Fig. 1b, where the sender's probability of choosing a certain message gradually changes, and the receiver strategy assigns a positive probability to more than one state for each message. What characterizes vagueness in these models is thus a smooth and monotonous transition between parts of the state space where a message is clearly used and where it is clearly not. This means that, for some states in the middle of the state space, there is uncertainty as to which message will be used by the sender, whereas for states in the extremes this is as good as certain. For the receiver part, there is uncertainty as to which state will be picked for each message.

The interpretation of this uncertainty will be different depending on the interpretation of the model. If we see it as an explanatory model of how two agents play the game, we can see it as randomization. If we interpret the model as descriptive, it simply represents expected behavior in a manner agnostic to the underlying mechanisms. A third option is to see probabilities as capturing relative numbers in a population of agents. For example, if the sender strategy assigns a probability of 0.4 to the event of message m being sent for state t, this would mean that 40% of the population uses that message for that state. This latter option leaves open the possible interpretation that each agent commands a crisp language and vagueness is only a population-level effect. However, given the level of abstraction of the description so far, none of this is necessarily implied by the model. Our preferred stance is to see the model as descriptive, but we do not wish to focus on that debate here. The question we want to address is: what additional modifications to sim-max games are sufficient for optimal languages to be more like Fig. 1b, rather than like Fig. 1a?

Franke, Jäger, and van Rooij (2011) make two suggestions as to how vagueness in a signaling game can be boundedly rational, i.e. how vagueness could arise as a consequence of cost-saving limitations in the cognitive capacities of instrumentally rational agents. The first proposal is called limited memory fictitious play (LMF) and models agents playing a sim-max game where their ability to recall past interactions with others is limited to a certain number. For a given interaction, each agent uses her limited memory of the other agent's past behavior to estimate the other's strategy, and plays an instrumentally rational best response to that strategy. In order to study the evolution of strategies in repeated interaction, the authors model several individual agents in actual play. What they observe is the emergence of vague signaling at the level of the population, i.e. population averages of individual strategies exhibit the hallmarks of a vague language. This is because each agent recalls a different history of previous play and so holds slightly different beliefs about language use. However, each agent still commands a crisp language, which is an inadequate feature of the model if the intention is to capture how vagueness presents itself in human languages.

In order to overcome this limitation, Franke, Jäger, and van Rooij make another proposal using the notion of a quantal response equilibrium (QRE). The idea, inspired by experimental psychology, is to model the choice of best response as stochastic rather than deterministic.[4] A prominent explanation for such soft-max or quantal choice behavior is that agents make small mistakes in the calculation of their expected utilities (Train 2009). They still choose the option with the highest expected utility, but each assessment of the expected utilities is noise perturbed. This, in turn, may actually be boundedly rational since the calculation of expected utilities relies on assessing stochastic uncertainty, which in turn may be costly to calculate precisely. Choice based on a few samples from a distribution can be optimal if taking more samples or other means of better approximating probabilistic beliefs is resource costly (e.g. Vul et al. 2014; Sanborn and Chater 2016). The degree to which agents tremble in the calculation of expected utilities and therefore deviate from the instrumentally rational behavior can be characterized by a parameter. Franke, Jäger, and van Rooij find that for low values of this parameter, only babbling equilibria are possible, where sender and receiver simply randomize, respectively, message and interpretation choice uniformly. Above a certain value of the parameter, other equilibria of the kind described in the beginning of this section arise, where agents communicate successfully, though not perfectly, using fuzzy strategies similar to those depicted in Fig. 1b. However, it is not clear whether soft-max choices capture the right stochastic trembles in decision making as they would arise under natural sources of uncertainty about the context (see Franke and Correia 2018).

O'Connor (2014) proposes a way in which vagueness could be expected to evolve as a side-effect of a particular type of learning process. She studies sim-max games driven not by rational choice dynamics, but by generalized reinforcement learning (GRL), a variant of Herrnstein reinforcement learning (HRL) (Roth and Erev 1995). In HRL, agents learn to play a signaling game by strengthening particular choices (of messages for the sender, of responses for the receiver) proportionally to how successful those choices prove to be in an interaction. O'Connor's proposal is to model generalization as the propagation of reinforcement to nearby states, where "nearby" is defined in terms of distance in state space. For example, if a sender was successful in using message m for state t, she will not only positively reinforce that choice of message for t, but also for states similar to t. This is done to a degree that is proportional to the similarity between t and other states. The dynamics gives rise to vague signaling of the kind we are looking for.

Although there is a close relationship between reinforcement learning and population-level dynamics (Börgers and Sarin 1997; Beggs 2005), O'Connor's GRL is, on the face of it, an account of learning between individual players. Also, we need further justification for linking generalization in reinforcement so closely to the underlying payoff function of the sim-max game. Why should agents evolve to generalize in exactly the right way? O'Connor (2015) suggests that, despite a language with vagueness having lower expected utility than a language without it, the learning mechanism that induces vagueness does have evolutionary advantages: it achieves higher payoffs in a shorter period of time. From a global point of view, learning speed can be an advantage. Imagine an initial population of agents with random strategies,

[4]Probabilistic choice rules are also the source of vagueness in recent accounts by Lassiter and Goodman (2017) and Qing and Franke (2014).

some using GRL and others using classic HRL. Although the latter type of agent can hypothetically develop a precise and more efficient signaling system, agents using GRL could coordinate on vague signaling strategies with high (though not optimal) expected utility sooner than agents using HRL. In such a scenario, they could drive the other agents to extinction before the latter had time to achieve coordination and reap the benefits of a more precise signaling system.

A similar finding is made by Franke and Correia (2018) when studying a variant of the replicator dynamics in which individual agents do not have the ability to generalize but simply make perceptual mistakes. In a scenario where agents learn by imitation, if one assumes that they do not have perfect perception, there will always be the possibility for senders to confuse states and thus learn associations with messages that are different than the ones observed, and for receivers to similarly mix up responses. Furthermore, it seems reasonable to assume that this confusability is proportional to state similarity, i.e. that the more similar two states are, the more likely it is that they will be mistaken for each other. Incorporating these considerations into a derivation of the replicator dynamics based on imitation processes, Franke and Correia develop a variant of the dynamic that also induces vague signal use of the kind we expect here. The consequence is very similar to that of the GRL model discussed above, in that the way the behavior for a given state is updated takes into account the behavior of similar states, proportionally to their similarity. Given the known relation between reinforcement learning and the replicator dynamics (Beggs 2005), it is actually quite plausible that the two are tightly related (although this would need to be formally demonstrated). The account is, furthermore, interpretable at a lower level of rationality.

The motivation underlying this model of vague signaling is still one of inevitability. A vague strategy is not claimed to have higher expected utility than a crisp one. However, the authors observe an effect similar to that pointed out by O'Connor: signaling converges faster and more often in scenarios where there is some degree of state confusability. Furthermore, they observe one additional potentially beneficial property. Running several rounds of simulation for each parameter set, they measure for each group of results how close resulting strategies are to each other, and how they would fare playing against one another. The results show that the within group distance between strategies becomes smaller with growing confusability, and the within group expected utility is actually higher for strategies evolving under a certain degree of state confusion. Thus, some amount of uncertainty seems to promote more homogeneous populations of signalers that are better at achieving cooperation within a group. What is left to show, as it was with O'Connor's GRL approach, is whether the potential payoff advantages that were observed in simulations actually suffice to promote vague language use in an encompassing model of multi-level selection. The following section motivates the need for taking a more ecological approach to the evolution of vagueness, before Sect. 5 gives a concrete model.

4 The Ecology of Vagueness

Arguments of the kind presented by Lipman (2009), that vague signal use is suboptimal when compared to crisp use, work under various assumptions. Part of the picture formed by these assumptions is a highly idealized conception of the agents involved and of the context in which they develop and use signals. These idealizations probably originate, via game theory, from the conception of rationality of traditional theoretical economics. Herbert A. Simon describes this picture as follows:

> Traditional economic theory postulates an "economic man," who, in the course of being "economic" is also"rational." This man is assumed to have knowledge of the relevant aspects of his environment which, if not absolutely complete, is at least impressively clear and voluminous. He is assumed also to have a well-organized and stable system of preferences, and a skill in computation that enables him to calculate, for the alternative courses of action that are available to him, which of these will permit him to reach the highest attainable point on his preference scale (Simon 1955, pp. 99).

Both Simon and Lipman call for this picture to be revised, and this is what the proposals surveyed here all do. In order to account for vagueness in natural language in the context of these models, they peel away from this idealized picture and bring some of these assumptions down to earth. In the process, they point to ways in which we, as language learners and language users, are finite beings finding ways to cope with a highly complex and dynamic environment:

1. Our existence is temporally finite; language does not have an infinite amount of time to evolve, nor can it take an infinite time to be learned. The faster a language can start being useful, the better;
2. Language learning through experience has to rely on a limited number of observations. Not only is the state space typically much larger than one can survey in sufficient time, it is even potentially infinite and constantly changing;
3. A corollary of the former is that there will always be heterogeneity in a population of language learners, at the very least in their prior experience, since each agent will have relied on a different set of observations. Furthermore, this information is not directly or fully accessible to others.

All of these observations support the weakening of modeling assumptions. The research surveyed here shows us some examples of assumptions which, when weakened, make vague signal use a natural outcome of certain dynamics. But it gives us even more. It suggests ways in which the mechanisms that lead to vagueness can have positive effects that are extremely important in the context of the points just enumerated. We learned from O'Connor (2014) and Franke and Correia (2018) that vague languages are quicker to converge and adapt, which is valuable given the finite and dynamic character of our experience (point 1). O'Connor (2014) also showed how generalization, an invaluable feature of any procedure for learning from a limited number of observations (point 2), leads to vagueness. We also learned from Franke and Correia (2018) that state confusability, a mechanism that leads to vague signal use, can have a homogenizing effect on vocabularies, potentially compensating for the heterogeneity of agents' experiences (point 3).

What do these observations tell us about rationality? GRL (O'Connor 2014) and the work of Franke and Correia (2018) both assume a picture of agents with a basic level of instrumental rationality, possibly limited awareness of the game, and a lack of strategic capabilities, adapting their behavior with only short-term gains in sight. These approaches introduce constraints on agent behavior or information processing that prevent the evolution of crisp signal use. But a crisp language would still have a higher expected utility than the evolved strategies. Agents in those models seem to be only as rational as the modelers allow them to be. Despite the plausibility of the mechanisms proposed (limited memory, imprecise calculation of expected utilities, generalization, state confusability), the results of these models feel somewhat bittersweet because of the hypothetical possibility of an ideal strategy, seemingly barred from the agents in an artificial manner. Couldn't a more rational agent evolve and drive the system into crisp signal use? Aren't we, human beings, that kind of agent?

Perhaps a deeper understanding of vagueness and the reasons for its pervasiveness in natural language are to be found only when we broaden the scope of the models employed. All the models discussed so far explore evolutionary dynamics for one homogeneous population playing one game. Different types of agents and different game setups are considered, but each of these different possible scenarios is tested separately. We see at least two ways in which one could embrace a more ecological perspective. The first is to think about meaning and vagueness from a more Wittgensteinian perspective. One can see each signaling game as embodying a particular language game. In Wittgenstein's picture of language, however, we do not play only one language game over the course of our existence; there is a plurality of them and which one an agent is engaged in at a particular moment is never clearly identified, neither are the exact benefits one might gain by choosing a certain behavior over another. These are furthermore not fixed in time; old language games fall out of fashion or stop being useful, and new ones emerge all the time (see Wittgenstein 1953, in particular §23).

We can look for rationality at several levels in this pluralistic picture. First, as before, there is the actual behavior of a single agent in each actualized language game. As mentioned above in connection the soft-max choice function used by Franke, Jäger, and van Rooij (2011), behavior that strictly maximizes expected utility under uncertainty may be resource heavy, so it might be compatible with local strategic rationality that agents' production choices are stochastic. Second, if we look at behavior across many game types and contexts, there is also the level of an agent's internal theory of how words and phrases are likely to be used (or even normatively: how messages should be used), conditional on a given context. Notice that a single agent's rational beliefs about linguistic practices or linguistic meaning may well have to reflect the actual stochasticity: under natural assumptions about information loss, the best belief for prediction matches the actual distribution in the real world (e.g. Vehtari and Ojanen 2012). In sum, both at the level of behavior and at the level of beliefs about use or meaning, we should expect to find vagueness. Still, despite vagueness, there does not seem to be anything fundamentally missing or conceptually incoherent in a naturalistic, rationality- or optimality-driven explanation of what each agent is doing or what each agent believes about language, use and meaning.

Another way to go beyond locality is to work with more heterogeneous population models. The mechanisms that lead to vague signal use, as O'Connor (2014) and Franke and Correia (2018) stress, have the aforementioned important advantages of faster speed of convergence, higher flexibility, and homogenization. The argument goes that these side-effects, by temporarily enabling a higher expected utility, could allow a population using some generalization (or affected by some imprecision) to take over. However, despite its intuitiveness, the argument is based on comparing isolated runs of different dynamics. The models do not allow the hypothesis to be tested, because they do not accommodate different populations evolving together. In the remainder of the paper we propose a way to do this based on the model of Franke and Correia (2018). We introduce two variants of a multi-population model of the imprecise imitation dynamics, where populations characterized by different imprecision values interact and evolve together. Using this model, we can better see under which conditions the hypothesized advantages of some imprecision can lead to the evolution of vagueness.

5 Two Multi-population Models of Imprecise Imitation

We build upon the model of Franke and Correia (2018), adding to the imprecise imitation dynamics support for multiple populations with different imprecision values evolving together. The resulting evolutionary dynamic has two layers: first, signaling strategies change by imprecise imitation following the model of Franke and Correia; second, there is evolutionary selection at the level of the imprecision values themselves. More concretely, imagine agents who are born with varying perceptual abilities, subsequently learn a signaling strategy by imitation of other agents (either within or across populations), and depending on the success of the strategies they develop, are more or less likely to survive and reproduce. In such a setup, evolutionary processes will not only change the distribution of strategies within a population, where a population is identified by a shared level of imprecision in imitation, they will also promote those levels of imprecision which result in the development of more successful signal use.

This is closely related to the ideas behind theories of kin selection (Hamilton 1964) and multi-level (or group) selection (Wilson 1975). These theories build upon the hypothesis that selection acts not only to directly favor genes that result in behaviors that benefit individuals, but also to indirectly favor genes that lead to behaviors that benefit either genotypically (kin) or socially (group) related individuals (see Okasha 2006, for more details). In our model there is a similar structure: there is a process of selection of behavior within each population, and levels of imprecision are selected across populations based on the strategies they give rise to. We believe there is an important difference with group and kin selection models in that the two selection processes in our model act on different entities: the inner shaping signaling behavior, and the outer selecting levels of imprecision. In any case, we intend our model to be descriptive rather than causal or explanatory. That means that we do not want to commit to seeing populations either as kin or as social groups, and use them as merely descriptive abstractions that allow us to capture the hypothetical impact of indirect selection processes.[5]

In the following, we lay down the formal details of our model. Let's start by defining A to be the set of imprecision values considered. For each value $\alpha \in A$, the proportion of its population is given by $P(\alpha)$, such that $\sum_{\alpha \in A} P(\alpha) = 1$. Each population has its own sender and receiver strategies, represented as σ^α and ρ^α. The probability that a given agent with imprecision α (or of type α) observes t_o when the actual state is t_a is given by $P_o^\alpha(t_o|t_a)$. If the same agent intends to realize t_i, the probability that she actually realizes t_r instead is given by $P_r^\alpha(t_r|t_i)$. Following Franke and Correia (2018, p.26), we can then define the following values.

Probability that t_a is actual if t_o is observed by an agent of type α:

$$P_{\bar{o}}^\alpha(t_a|t_o) \propto P_a(t_a) P_o^\alpha(t_o|t_a)$$

Probability that a random sender of type α produces m when the actual state is t_a:

$$P_\sigma^\alpha(m|t_a) = \sum_{t_o} P_o^\alpha(t_o|t_a)\sigma^\alpha(m|t_o)$$

Probability that the actual state is t_a if a random sender of type α produced m:

$$P_{\bar{\sigma}}^\alpha(t_a|m) \propto P_a(t_a) P_\sigma^\alpha(m|t_a)$$

Probability that t_r is realized by a random receiver of type α in response to message m:

$$P_\rho^\alpha(t_r|m) = \sum_{t_i} P_r^\alpha(t_r|t_i)\rho^\alpha(t_i|m)$$

These formulations are merely parameterized versions of the single-population model. They encapsulate calculations that one can use to compute expected utilities and strategy update steps for each type. The latter, however, depend on the types of interaction that we imagine occurring between populations. In the following sections, we consider two different possibilities.

5.1 Tight Population Interaction

In a multi-population model with tight interaction between populations, each agent plays with, observes, and potentially imitates any other agent, regardless of their type. This has an impact on the expected utilities of sender and receiver strategies of

[5] We make this note because of the heated debate between the two theories. See Kohn (2008), and Kramer and Meunier (2016) for more details.

each type, and on the update steps for those strategies. Let's start with the expected utilities. For a sender of type α, the expected utility of its strategy σ^α against all other receiver strategies $\rho^\star$, is given by:

$$\mathrm{EU}^\alpha_\sigma(m, t_o, \rho^\star) = \sum_{t_a} P^\alpha_{\bar{o}}(t_a|t_o) \sum_{\alpha' \in A} P(\alpha') \sum_{t_r} P^{\alpha'}_\rho(t_r|m) U(t_a, t_r)$$

and, for a receiver of type α, the expected utility of its strategy ρ^α against all other sender strategies $\sigma^\star$, is given by:

$$\mathrm{EU}^\alpha_\rho(t_i, m, \sigma^\star) = \sum_{\alpha' \in A} P(\alpha') \sum_{t_a} P^{\alpha'}_{\bar{\sigma}}(t_a|m) \sum_{t_r} P^\alpha_r(t_r|t_i) U(t_a, t_r)$$

Expected utilities thus take into account the existence of strategies of other types, and weigh the relevance of each type α' according to its relative proportion $P(\alpha')$.

Another important value to calculate has to do with the types that agents observe and imitate. In a model with tight interaction, we imagine this occurring across populations. Therefore, we can define the probability that a sender of type α observes a randomly sampled agent play message m for observed state t_o as:

$$P^\alpha_o(m|t_o) = \sum_{t_a} P^\alpha_{\bar{o}}(t_a|t_o) \sum_{\alpha' \in A} P(\alpha') P^{\alpha'}_\sigma(m|t_a)$$

and the probability that a receiver of type α observes a randomly sampled agent choose interpretation t_o given message m as:

$$P^\alpha_o(t_o|m) = \sum_{t_r} P^\alpha_o(t_o|t_r) \sum_{\alpha' \in A} P(\alpha') P^{\alpha'}_\rho(t_r|m)$$

Again, these calculations incorporate the probabilities that the imitating agent might observe the behavior of an agent of another type α', weighed by its relative proportion. Finally, the update step for a sender strategy of type α at time instant $i+1$ is given by:

$$\check{\sigma}^\alpha_{i+1}(m|t) \propto P^\alpha_o(m|t) \mathrm{EU}^\alpha_{\sigma_i}(m, t, \rho^\star_i)$$

and similarly for a receiver strategy of type α by:

$$\check{\rho}^\alpha_{i+1}(t|m) \propto P^\alpha_o(t|m) \mathrm{EU}^\alpha_{\rho_i}(t, m, \sigma^\star_i)$$

We here use $\check{\sigma}$ and $\check{\rho}$ since there is still an additional adjustment to these values to be calculated before we get the final strategies σ and ρ.

These formulations cover the evolution of the particular strategies of each type. We can think of this as the level of cultural evolution: agents are born with a certain level of imprecision and adopt strategies based on the behavior of others. Alongside this process, we can imagine another level of selection, where agents die and new agents are born. More successful agents have a higher likelihood of surviving and reproducing, giving rise to more agents with their level of imprecision. Levels of imprecision are thus subject to an evolutionary dynamic that is indirectly influenced by the cultural dynamic. Importantly, only the level of imprecision is passed on to new generations under this dynamic, not the actual strategies developed by the agents at the cultural level.

We model this process by changing the proportion of each type $P(\alpha)$ according to the replicator dynamic. A population of type α consists of agents employing both a sender and a receiver strategy, thus the overall fitness of the population must include the expected utilities of both. We could imagine this process happening at a different speed to the cultural process, in which case we could have different time scales. For the sake of simplicity, we choose to have them both happen at each time step, but calculate the changes occurring at this level of selection after the calculation for the other level. We define the proportion of type α at time step $i + 1$ as:

$$P_{i+1}(\alpha) \propto P_i(\alpha)(\mathrm{EU}^{\alpha}_{\sigma_{i+1}}(m, t, \rho^{\star}_i) + \mathrm{EU}^{\alpha}_{\rho_{i+1}}(t, m, \sigma^{\star}_i))$$

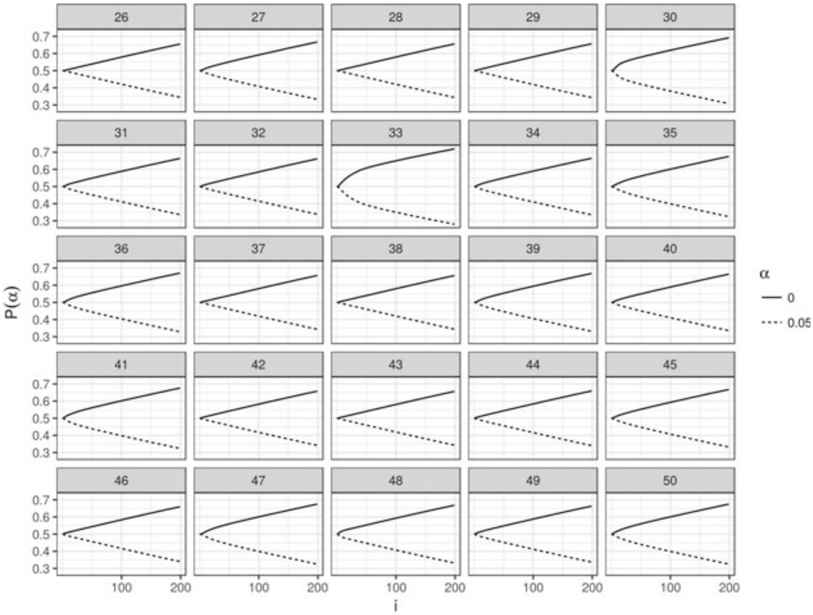
In order to additionally account for the fact that strategies are not passed on to new generations, we mix the evolved strategies of a certain type σ^{α} and ρ^{α} with new random strategies $\tilde{\sigma}^{\alpha}$ and $\tilde{\rho}^{\alpha}$ (generated at each time step). The idea is that, at each time step, a certain percentage of each population will consist of "newborn" agents, i.e. agents that haven't yet had time to evolve their strategies. We define a parameter γ that quantifies this percentage, or as we can also call it the birth rate, which we consider to be the same for every population. This mixing finally defines the strategies for time step $i + 1$ and can be described in the following formulas:

$$\sigma^{\alpha}_{i+1}(m|t) = (1 - \gamma)\check{\sigma}^{\alpha}_{i+1}(m|t) + \gamma\tilde{\sigma}^{\alpha}_{i+1}(m|t)$$

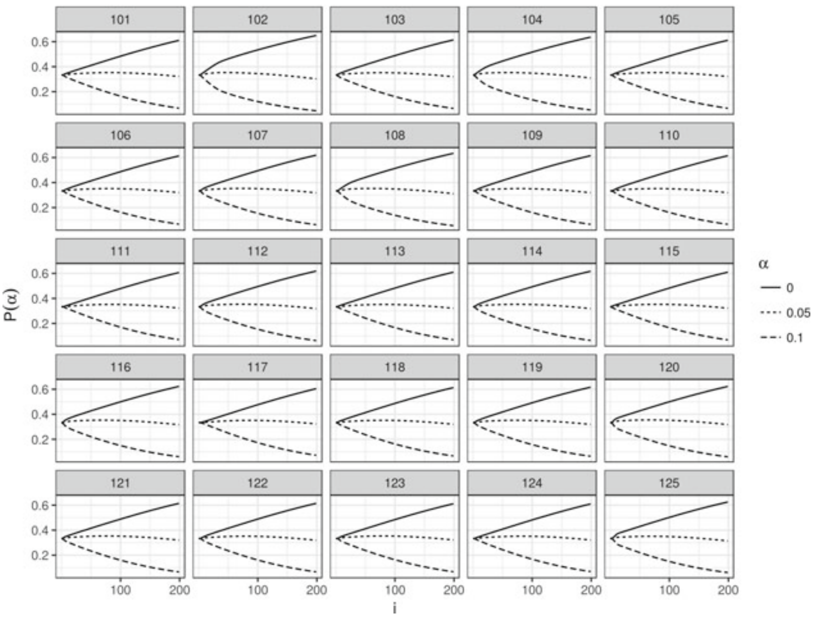
$$\rho^{\alpha}_{i+1}(t|m) = (1 - \gamma)\check{\rho}^{\alpha}_{i+1}(t|m) + \gamma\tilde{\rho}^{\alpha}_{i+1}(t|m)$$

Simulation results We performed 25 simulation runs of this model for each of three population scenarios: only one population with $\alpha = 0$ (for reference), two populations with $\alpha \in \{0, 0.05\}$, and three populations with $\alpha \in \{0, 0.05, 0.1\}$. For each scenario, starting proportions $P(\alpha)$ were equal for each value of α. Given the observations by Franke and Correia (2018) that state space size and tolerance parameter β do not result in important qualitative difference, we fixed these values at $n_S = 30$ and $\beta = 0.1$. We also used a fixed uniform distribution for the priors and a message space with 2 messages. Each type started with its own randomly generated strategy. Regarding the duration of the simulations, due to the mixing in of new individuals into the population (birth rate was fixed at $\gamma = 0.05$), the convergence criteria is no longer applicable because each strategy is randomly perturbed at each time step. Therefore, all simulation runs were stopped after 200 iterations.

The first thing to observe from the simulation results is that the population with no imprecision ($\alpha = 0$) dominated the other populations in every run. In Fig. 2 we plot the evolution of the proportion of each population in the two-population and



(a) Two-population scenario.



(b) Three-population scenario.

Fig. 2 Evolution of population proportions through time for each simulation trial of the tight interaction model. Numbers on top of each plot identify each trial three-population scenarios for all trials. As the plot shows, the population with no imprecision steadily increased its proportion against the others in every trial. In the three-population scenario, the proportion of type $\alpha = 0.05$ sees a slight increase in the beginning of the simulation, but inexorably starts a downward trend. These observations go against the expectation of Franke and Correia (2018) that faster convergence to a convex strategy by populations with a certain level of imprecision could give them a temporary advantage to take over and eliminate other types. The reason for this is interesting in itself. What happens is that, because of the tight interaction between populations, the strategies of each type evolve in close tandem with each other. One of the consequences of this is that the population with no imprecision reaches convexity faster than it would on its own because of the interaction with the populations with imprecision. We can see this effect by looking into the percentage of trials with convex sender strategies $\alpha = 0$ at a given iteration, for each scenario, and comparing the three scenarios: populations with no imprecision evolving alone, two populations ($\alpha \in \{0, 0.05\}$), and three populations ($\alpha \in \{0, 0.05, 0.1\}$). This is plotted in Fig. 3a. What we see is more trials reaching convexity earlier for the multi-population scenarios when compared with the single-population scenario. This effect precludes the hypothesized temporary advantage of imprecision manifesting itself, but it can be seen as a positive influence of the population(s) with imprecision on the population with no imprecision.

The flip side of this tight connection between populations is that strategies evolved by populations with no imprecision are also more vague (in the sense defined in Sect. 3). This is visible by looking at mean entropy values $\alpha = 0$, namely sender strategy entropy, plotted in Fig. 3b, and receiver strategy entropy, plotted in Fig. 3c. The values for the population with no imprecision are clearly higher in the scenarios where it evolves together with populations with imprecision than in the scenario where it evolves on its own. Given the trends in population proportions, one expects this to eventually be eliminated when the population with no imprecision finally takes over the others, but it is interesting to observe that while populations with vague strategies persist, the population with no imprecision takes much longer to evolve a fully crisp strategy.

5.2 Loose Population Interaction

We can also imagine a scenario where populations interact more loosely. Namely, the dynamic we want to model here is one where agents of a certain type α imitate and learn only from other agents of the same type, but nevertheless use their signaling strategies with agents from all types. In order to capture this, we need to make changes to some calculations. Any formula that is not redefined in this section should be assumed to remain the same. First, if agents learn only within their population, the imitation dynamic needs to consider only the expected utility against agents of that population. We thus define the following expected utilities for a sender of type α:

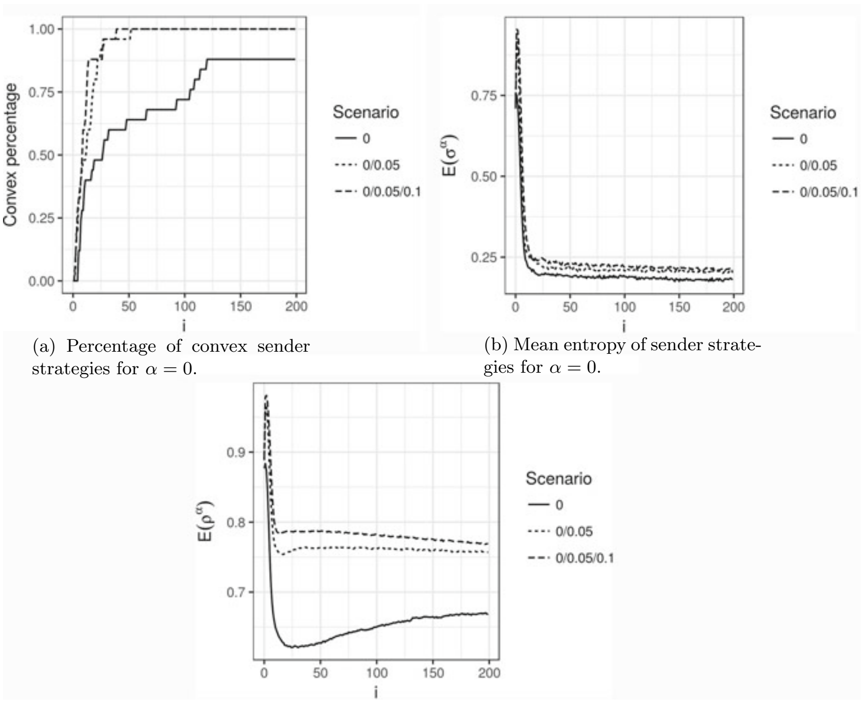
$$\mathrm{EU}^{\alpha}_{\sigma}(m, t_o, \rho^{\alpha}) = \sum_{t_a} P^{\alpha}_{\bar{o}}(t_a|t_o) \sum_{t_r} P^{\alpha}_{\rho}(t_r|m) U(t_a, t_r)$$

and for a receiver of type α:

$$\mathrm{EU}^{\alpha}_{\rho}(t_i, m, \sigma^{\alpha}) = \sum_{t_a} P^{\alpha}_{\bar{\sigma}}(t_a|m) \sum_{t_r} P^{\alpha}_{r}(t_r|t_i) U(t_a, t_r)$$

The main difference from the previous model is that these expected utilities are not calculated against all populations ($\sigma^{\star}$ and $\rho^{\star}$) but only against the agent's own type (σ^{α} and ρ^{α}). This also implies that population proportions do not play a role at this level of selection.

Regarding the imitation process, we can redefine the probability that a sender of type α observes a randomly sampled agent play message m for observed state t_o as:



(a) Percentage of convex sender strategies for $\alpha = 0$.

(b) Mean entropy of sender strategies for $\alpha = 0$.

(c) Mean entropy of receiver strategies for $\alpha = 0$.

Fig. 3 Development of some metrics through time for the $\alpha = 0$ population in each of three scenarios: evolving alone ('0'), with an $\alpha = 0.05$ population ('0/0.05'), and with an additional $\alpha = 0.1$ population ('0/0.05/0.1')

$$P_o^\alpha(m|t_o) = \sum_{t_a} P_{\bar{o}}^\alpha(t_a|t_o) P_\sigma^\alpha(m|t_a)$$

and the probability that a receiver of type α observes a randomly sampled agent choose interpretation t_o given message m as:

$$P_o^\alpha(t_o|m) = \sum_{t_r} P_o^\alpha(t_o|t_r) P_\rho^\alpha(t_r|m)$$

Again, the main difference is that agents make observations within their own population, so to model a randomly sampled agent one needs only to take into account agents of the same type. Population proportions again do not play a role in these calculations. Based on these formulas, we can define the update step for a sender strategy of type α at time instant $i+1$ as:

$$\check{\sigma}_{i+1}^\alpha(m|t) \propto P_o^\alpha(m|t)\mathrm{EU}_{\sigma_i}^\alpha(m, t, \rho_i^\alpha)$$

and similarly for a receiver strategy of type α as:

$$\check{\rho}_{i+1}^\alpha(t|m) \propto P_o^\alpha(t|m)\mathrm{EU}_{\rho_i}^\alpha(t, m, \sigma_i^\alpha)$$

Note that, because imitation and learning occur only within populations, these formulations are essentially the same as for the single-population model of Franke and Correia (2018), only parameterized by type α.

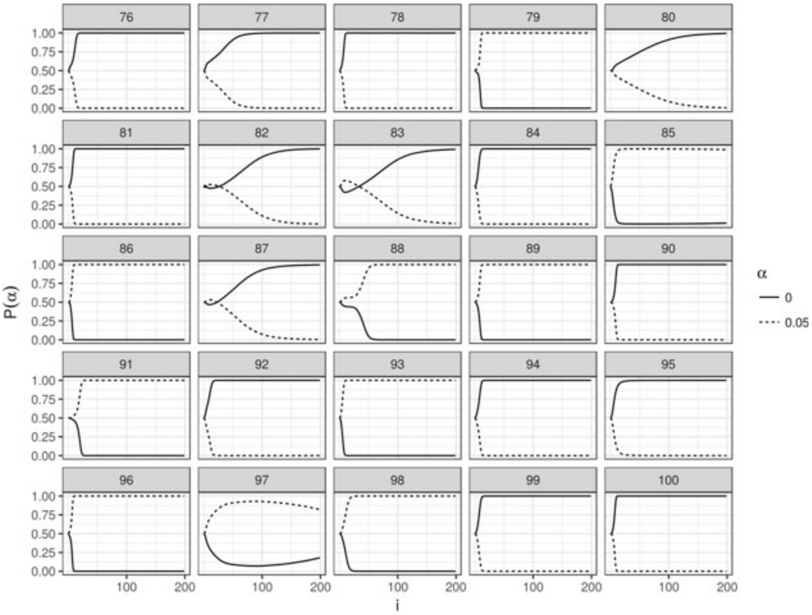
The selection process between different populations still follows the same motivation as before: agents of a certain type that evolve successful strategies (with respect to other types) will be more likely to survive and reproduce, benefiting the proportion of their population. The formulation of the dynamic for the proportion of each type $P(\alpha)$ thus stays the same. In order to avoid confusion, we want to stress that this means that these calculations rely on the definitions of expected utility across populations (i.e. $\mathrm{EU}_\sigma^\alpha(m, t_o, \rho^\star)$ and $\mathrm{EU}_\rho^\alpha(t_i, m, \sigma^\star)$) and not the newly introduced $\mathrm{EU}_\sigma^\alpha(m, t_o, \rho^\alpha)$ and $\mathrm{EU}_\rho^\alpha(t_i, m, \sigma^\alpha)$.

Simulation results We ran the same number of simulation trials under the same conditions as for the model with tight population interaction. In Fig. 4 we plot the evolution of population proportions for all trials of the multi-population scenarios. The first thing to observe is that proportions evolve faster than in the model with tight interaction. Whereas in the latter no given population ever reached much more than 70% proportion, in this model we see that many trials resulted in one population fully dominating the others. In the two-population scenario, some population reached at least 99% in 24 out of 25 trials. In the three-population scenario, this happened in 18 out of 25 trials. More interestingly for our investigation, some trials actually resulted in the population with $\alpha = 0$ being dominated. For the two-population scenario, $\alpha = 0.05$ fully reached 100% in 8 trials (79, 86, 88, 89, 91, 93, 96, 98), a point from which it is technically impossible for the other population to recover. In most cases the dominating population gains its ground from the start, but in 5 cases we see a temporary advantage of $\alpha = 0.05$ that is then lost to the other population. In one interesting trial (85), $\alpha = 0.05$ reached 99.87% only to then steadily start losing ground to $\alpha = 0$. For the three-population scenario, $\alpha = 0$ was reduced to 0% in 4 trials (152, 157, 159, and 175). In all of those cases, $\alpha = 0.05$ clearly has the upper hand over $\alpha = 0.1$, despite a temporary advantage of the latter in some trials.

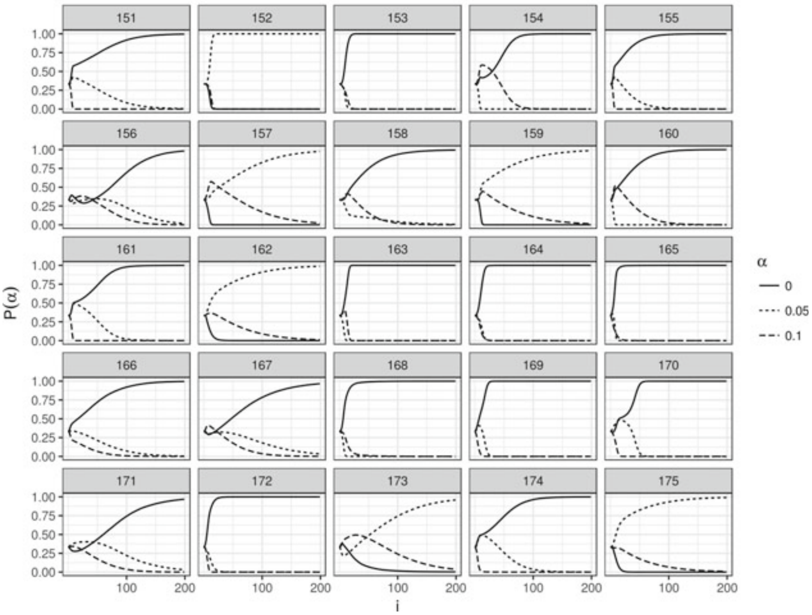
Because of the loose interaction between populations, different types can now evolve separately. One consequence of this is that populations with a certain level of imprecision again reach convexity faster than those without imprecision. In Fig. 5 we plot, for each trial, the first iteration when each type reached convexity. What we see is that, even though $\alpha = 0$ usually reaches convexity later, this is not always the case. This has certainly to do with the initial conditions of each trial, since the randomly generated strategies can simply by chance be more favourable to reaching convexity. More importantly, we also see that reaching convexity sooner is not a sufficient condition for achieving population dominance. There were many trials where another type reached convexity much sooner than $\alpha = 0$ but its population was completely dominated nevertheless (e.g. 76, 84, 92, and 99 in the two-population scenario; 156, 165, 168, and 172 in the three-population scenario). It also does not seem to be fully necessary, given a few examples where $\alpha = 0$ reached convexity early on and another type ended up dominating (89 in the two-population scenario and 152 in the three-population scenario).

Another consequence of the populations evolving separately in this model is that they do not necessarily evolve towards the same equilibrium. In a sim-max game such as the one set up here, there are only two stable equilibria. These are the two Voronoi languages of the kind shown in Fig. 1: one where the first message is used for the first half of the state space, and the other where the second message is used for this region. In the multi-population model with loose interaction, each population evolves independently towards one of these two equilibria. An important factor determining which language strategies converge to is the random initial population configuration. The populations are, however, not fully independent, since the process of selection of the level of imprecision relies on the expected utility of one population playing against itself and the others. And this is important because strategies in one equilibrium get the lowest payoff possible against strategies in the other equilibrium. Now, if two populations evolve towards different equilibria, whatever advantage one population has playing against itself could trigger a runaway effect by causing an increase in its proportion, which in turn will increase the population's relative expected utility, potentially increasing their proportion further in the next round, and so forth. In this case, one would expect that faster convergence towards convexity would be especially important for a population's success.

In the two-population scenario, of the 8 trials where $\alpha = 0$ was reduced to 0%, all but two (88, 98) ended with the two populations each close to different equilibria. In the three-population scenario, the population with $\alpha = 0$ evolved towards the same equilibrium as the other two in 1 of the trials where it was eliminated. In the other 3 trials, in 2 it evolved to a different equilibrium than the other two, and in 1 it evolved towards the same equilibrium as $\alpha = 0.05$ (but different to the one $\alpha = 0.1$ evolved

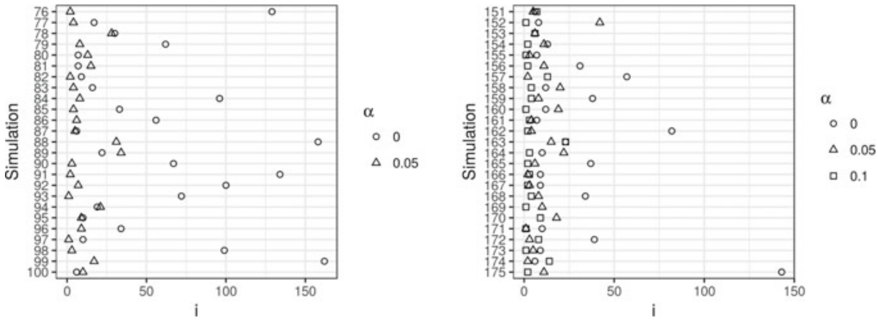


(a) Two-population scenario.



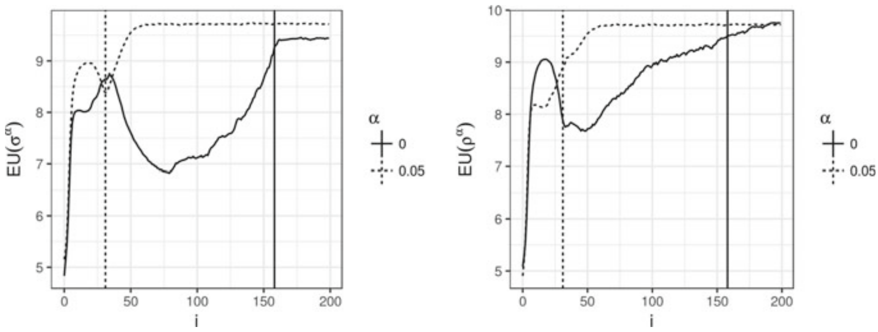
(b) Three-population scenario.

Fig. 4 Evolution of population proportions through time for each simulation trial of the loose interaction model. Numbers on top of each plot identify each trial



(a) Two populations scenario. (b) Three populations scenario.

Fig. 5 First iteration with convex sender strategies for each simulation trial of the loose interaction model



(a) Sender expected utility. (b) Receiver expected utility.

Fig. 6 Expected utilities for trial 88 of the loose interaction model. Vertical lines demarcate, for each type, the iteration where convexity was achieved

towards). Again, we do not find a clear case for this being a decisive factor in the success of populations with some imprecision. The same goes for the impact of an initial advantage in expected utility creating the runaway effect we just mentioned. Despite none of these three factors (including reaching convexity sooner) sentencing the demise of $\alpha = 0$ with certainty, they do seem to conspire together to bring it about. In most trials where the type was eliminated, in both the two- and three-population scenarios, $\alpha = 0$ ended up evolving towards a different equilibrium than the other types, and either had an initial disadvantage in expected utility, or reached convexity much later. These factors can thus be seen as indicators at best, and the story of how a population with imprecision ends up dominating one without it seems to be more complicated than expected.

This is not surprising, since we are facing a complex dynamical system with various interacting components (sender and receiver strategies and multiple populations). Just as an example of this, in Fig. 6 we plot the expected utilities of sender and receiver strategies for both populations of trial 88. The curves up to where $\alpha = 0.05$ achieved convexity seem to suggest that the sender strategy of that type was evolving towards the same equilibrium as the receiver strategy of the other. The moment where the population reaches convexity marks the point where the sender strategy of $\alpha = 0.05$ aligns with its type's receiver strategy and this coincides with the moment when the population seems to gain real traction over the other type (see again the plot for trial 88 in Fig. 5). In this particular case, reaching convexity seems to have made an important difference.

6 Conclusions

Vagueness presents a challenge to both procedural and instrumental pictures of rationality alike. In the context of game-theoretical models of language, this takes the form of a question about evolution: how can vagueness persist if it characterizes demonstrably less efficient communication? Most existing proposals in the literature attempt to explain this by considering agents with some degree of bounded rationality. Despite the many ways in which our rationality is inevitably limited, the argument for the pervasiveness of vagueness in natural language would be much stronger if one could also find associated advantages. In this paper, we argued that finding those might require us to go beyond a local notion of rationality, as two of the proposals reviewed here (O'Connor 2014; Franke and Correia 2018) suggest. We advocate moving towards an ecological approach, studying vagueness in more heterogeneous ecosystems. Language is part of a very complex system (Beckner et al. 2009) that involves many components interacting in often unpredictable ways. An ecology of vagueness would involve studying models where different populations can evolve and interact with each other, where different language games can be played between individuals, where the environment is uncertain and changing, or anything else that more closely approximates the real context of language evolution.

In light of this picture, we have proposed two variants of a concrete multi-population signaling model to test the hypothesis that certain features of imprecise imitation, like promoting faster convergence and regularity, could prove beneficial in contrast with full precision. Analysing these models did not provide us with a clear-cut answer, and revealed that the story is much more nuanced than initially expected. In a variant where populations with and without imprecision interact tightly, although precision always has the upper hand, vagueness seems to take a long time to be weeded out. When we let populations interact more loosely, we see a more complex pattern of outcomes. These include scenarios where imprecise imitation dominates over full precision, showing that strategies with vagueness can actually, under certain circumstances, be more successful. Bringing several populations together in a more complex ecosystem thus allowed us to not only spell out and test the original intuition, but also learn about unforeseen effects. These models thus serve as an example of how moving to a more global perspective on rationality can allow us to achieve a more detailed awareness of the complex interactions that might be involved in sustaining vagueness in natural language. They can be seen, we believe, as a first step towards an ecology of vagueness.

References

Barrett, J. A. (2009). The evolution of coding in signaling games. *Theory and Decision*, *67*(2), 223–237.

Barrett, J. A. (2010). Faithful description and the incommensurability of evolved languages. *Philosophical Studies*, *147*(1), 123–137.

Beckner, C., et al. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Beggs, A. (2005). On the convergence of reinforcement learning. *Journal of Economic Theory*, *122*(1), 1–36.

Benz, A., Jäger, G., & van Rooij, R. (Eds.). (2006). *Game Theory and Pragmatics*. Hampshire: Palgrave MacMillan.

Börgers, T., & Sarin, R. (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, *77*(1), 1–14.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Dawkins, R. (1978). *The selfish gene* (New ed.). London: Flamingo.

Fawcett, T. W., Hamblin, S., & Giraldeaub, L.-A. (2013). Exposing the behavioral gambit: The evolution of learning and decision rules. *Behavioral Ecology*, *24*(1), 2–11.

Franke, M., & Correia, J. P. (2018). Vagueness and imprecise imitation in signalling games. *The British Journal for the Philosophy of Science, 69*(4), 1037–1067. https://doi.org/10.1093/bjps/axx002.

Franke, M., Jäger, G., & van Rooij, R. (2011). Vagueness, signaling and bounded rationality. *New frontiers in artificial intelligence* (pp. 45–59). Berlin: Springer.

Franke, M., & Wagner, E. (2014). Game theory and the evolution of meaning. *Language and Linguistics Compass*, *8*(9), 359–372.

Galeazzi, P., & Franke, M. (2017). Smart representations: Rationality and evolution in a richer environment. *Philosophy of Science*, *84*(3), 544–573.

Gilboa, I., & Matsui, A. (1991). Social stability and equilibrium. *Econometrica: Journal of the Econometric Society*, 859–867.

Hagen, E. H. et al. (2012). Decision making: What can evolution do for us? In P. Hammerstein & J. R. Stevens (Eds.), *Evolution and the mechanisms of decision making* (pp. 97–126). Cambridge: MIT Press.

Hamilton, W. D. (1964). The genetical evolution of social behaviour I. *Journal of Theoretical Biology*, *7*(1), 1–16.

Hume, D.(1738/2005). *A treatise of human nature*. New York: Barnes & Noble.

Huttegger, S. (2014). How much rationality do we need to explain conventions? *Philosophy Compass*, *9*(1), 11–21.

Huttegger, S., et al. (2014). Some dynamics of signaling games. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(Suppl 3), 10873–10880.

Jäger, G. (2007). The evolution of convex categories. *Linguistics and Philosophy*, *30*(5), 551–564.

Jäger, G., Metzger, L. P., & Riedel, F. (2011). Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals. *Games and Economic Behavior*, *73*(2), 517–537.

Jäger, G., & van Rooij, R. (2007). Language structure: Psychological and social constraints. *Synthese*, *159*(1), 99–130.

Keefe, R., & Smith P. (1997). Introduction: theories of vagueness. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader* (pp. 1–57). Cambridge: MIT Press.

Kohn, M. (2008). Darwin 200: The needs of the many. *Nature News*, *456*(7220), 296–299.

Kramer, J. & Meunier, J. (2016). Kin and multilevel selection in social evolution: A never-ending controversy? In F1000Research 5.776.

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, *194*(10), 3801–3836.

Lewis, D. (1969/2011). *Convention: A philosophical study*. Oxford: Blackwell.

Lipman, B. L. (2009). *Why is language vague?* http://people.bu.edu/blipman/Papers/vague5.pdf.
Smith, J. M. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.
Mühlenbernd, R. (2011). Learning with neighbors: Emergence of convention in a society of learning agents. *Synthese*, *183*, 87–109.
O'Connor, C. (2014). The evolution of vagueness. *Erkenntnis*, *79*(4), 707–727.
O'Connor, C. (2015). Evolving to generalize: Trading precision for speed. *The British Journal for the Philosophy of Science, 68*(2), 309–410.
Okasha, S. (2006). *Evolution and the levels of selection*. Oxford: Oxford University Press.
Pagel, M. (2009). Human language as a culturally transmitted replicator. *Nature re-views genetics*, *10*, 405–415.
Qing, C. & Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In J. Grieser et al. (Eds.), Proceedings of SALT 44. (pp. 23–41).
Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, *8*(1), 164–212.
Russell, B. (1923). Vagueness. *Australasian Journal of Psychology and Philosophy*, *1*(2), 84–92.
Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893.
Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.
Simon, H. A. (1986). Rationality in psychology and economics. *Journal of Business*, S209–S224.
Skyrms, B. (1996). *Evolution of the social contract*. Cambridge: Cambridge University Press.
Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford: Oxford University Press.
Sorensen, Roy (2009). Sorites arguments. In Kim J., Sosa E., & Rosenkrantz G. S. (Eds.), *A companion to metaphysics* (2nd ed. pp. 565–566). New York: Wiley.
Spike, M., et al. (2017). Minimal requirements for the emergence of learned signaling. *Cognitive Science*, *41*(3), 623–658.
Taylor, P. D., & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Bioscience*, *40*(1–2), 145–156.
Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(16), 4530–4535.
Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.
Unger, P. (1979). There are no ordinary things. *Synthese*, *41*(2), 117–154.
Vanderschraaf, P. (1998). The informal game theory in Hume's account of convention. *Economics and Philosophy*, *14*(02), 215–247.
Vehtari, A., & Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, *6*, 142–228.
Vul, E., et al. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.
Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(1), 143–146.
Wittgenstein, L. (1953). Philosophical investigations. Oxford: Basil Blackwell.
Zollman, K. J. S., & Smead, R. (2010). Plasticity and language: An example of the Baldwin effect? *Philosophical Studies*, *147*(1), 7–21.

The Rationality of Vagueness

Igor Douven

Abstract Vagueness is often regarded as a kind of defect of our language or of our thinking. This paper portrays vagueness as the natural outcome of applying a number of rationality principles to the cognitive domain. Given our physical and cognitive makeup, and given the way the world is, applying those principles to conceptualization predicts not only the concepts that are actually in use, but also their vagueness, and how and when their vagueness manifests itself (insofar as the concepts are vague).

1 Introduction

Vagueness is often characterized in terms of the existence of so-called borderline cases. To say that a predicate is vague is to say that there are cases to which the predicate neither clearly applies nor clearly fails to apply. Such a situation of semantic indecision can easily appear unfortunate, and the study of vagueness may seem to some like studying a pathology of language. Too bad linguists and logicians were not around when our ancestors stumbled upon the possibility of verbal communication, tens of thousands of years ago; then we would not now be saddled with a language that more often than not resists formalization in the systems taught in introductory and even not so introductory courses on logic.

On reflection, however, one realizes that this view on the matter is too harsh. Linguists and logicians could have done little or nothing to enhance human discriminatory capacities, nor could they have made the world "gappier." Thus, researchers interested in vagueness tend to conceive of it as an unfortunate phenomenon, but also as one that is inevitable, given our physical and cognitive makeup and the way the world is.

In this paper, I take a less apologetic approach to vagueness. I propose that the occurrence of vagueness is not just excusable, or understandable, or inevitable; the "structure" of vagueness—by which I mean, which predicates are vague and where

I. Douven (✉)
Sciences, Normes, Décision (CNRS), Sorbonne University, Paris, France
e-mail: igor.douven@paris-sorbonne.fr

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*,
Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_7

their borderlines are found—can be seen as following from principles of rational design applied in the cognitive realm.

The design principles to be discussed actually concern *concepts*—the cognitive correlates of predicates—which in this paper are taken to be the primary bearers of vagueness. In a popular framework for modeling concepts, concepts reside in so-called similarity spaces (see below). The main claim is that if such spaces are optimally designed, then we should *expect* to see vagueness arise when and where, as a matter of empirical fact, it does arise.

Section 2 briefly describes the conceptual spaces framework, which is an increasingly popular format for theorizing about concepts and will serve as a background for the proposals made in this paper. I should say at the outset that, while popular, it is still unclear exactly how broadly applicable the framework really is. So far, it has been used with great success for the representation of perceptual concepts, and important progress has recently been made in applying the framework to more abstract—especially scientific—concepts. We should not be surprised, however, if it turns out that the framework is limited in terms of the types of concepts it can represent. I set this issue aside, however, for the remainder of this paper.

In Sect. 3, I outline my favorite theory of vagueness, which relies heavily on the machinery of the conceptual spaces framework. It relies just as heavily on a proposal made in Gärdenfors (2000), which gives pride of place to prototypes and so-called Voronoi tessellations for arriving at specific structures of spaces. I show how this proposal yields a formally precise account of vagueness, if we are willing to take on board a seemingly uncontroversial assumption about prototypes and slightly modify the technique of Voronoi tessellations.

Gärdenfors' proposal, and concomitantly the account of vagueness based upon it, leaves some important questions unanswered. Most notably, while Gärdenfors stresses from the outset that his proposal is meant to pertain to *natural* concepts—concepts that have, or could plausibly have, a role in our everyday thinking and theorizing, and that therefore we might care to name in our language—he admits that he has probably not managed to characterize the relevant notion of naturalness in a fully satisfactory way. And indeed, on his original account, too many concepts qualify as natural, including ones that pre-theoretically are clearly *not* natural.

Although not exactly presented in this way, Gärdenfors' attempt to differentiate natural concepts from non-natural ones appeals to a kind of rationality principle (or design principle). Natural concepts—the claim runs—are those that resulted from carving up a conceptual space in the most rational manner. As mentioned, this proposal does not quite do the job. But as argued in Douven and Gärdenfors (2018), this is not because it is a mistake to try to connect naturalness with rational design, but rather because not enough rationality principles are invoked to determine a carving-up of any given space. Douven and Gärdenfors list a number of rationality principles that—they argue—together might just suffice to yield carvings-up of conceptual spaces that result in truly natural concepts. Their claim, in slogan form, is that natural concepts are concepts represented by the cells of an optimally (i.e., most rationally) partitioned similarity space. This proposal is summarized in Sect. 4.

Douven and Gärdenfors (2018) explicitly state that they leave aside the issue of how to accommodate vagueness within their new proposal. That issue is taken up in Sect. 5 of this paper. Broadly speaking, the answer will be that the account of vagueness mentioned above is not only consistent with the new proposal, but that the two fit together very naturally, and can be seen as both being part of a three-step theory of conceptualization, with rationality playing a key role at every stage.

2 The Conceptual Spaces Framework

In the 1960s and 1970s, cognitive psychologists started using so-called dimensionality reduction techniques, such as principal component analysis and multi-dimensional scaling, to give low-dimensional representations of people's similarity judgments concerning (often) high-dimensional data (see, e.g., Krumhansl 1978; Borg and Groenen 2010). The outcomes of these procedures are commonly referred to as "similarity spaces."

To illustrate, consider the example of faces. Faces can vary along multiple dimensions: height, width, shape of the nose, eyes, eyebrows, color of the skin, and on and on. When two faces strike us as looking very similar, they probably differ very little along all the relevant dimensions, or at least along most of them. One would expect, then, that if we are to have a formal model that allows us to predict with some accuracy whether or not people will judge two given faces to be similar, the model must have separate parameters for all or most of the dimensions along which faces can vary.

Surprisingly, this turns out not to be the case. Low-dimensional models—so-called "face spaces"—have been developed that tend to yield precise predictions about people's similarity judgments of faces. Indeed there exist three-dimensional face spaces that have proven highly successful in this respect; see Valentine et al. (2016). What this means is that faces can be located in a three-dimensional space in such a way that measuring the distance in the space between any pair of faces gives a good indication of how similar to each other those faces will be deemed. Specifically, the prediction—which in the meantime has been supported by a wealth of evidence—is that the closer two faces are in the space, the more similar they will be judged to be.[1] The same is known to hold for colors, odors, sounds, tastes, various types of actions, and various types of shapes. Accordingly, we find in the scientific literature references to color spaces, odor spaces, taste spaces, vowel spaces, consonant spaces, action spaces, shape spaces, and others.

In the first instance, these are all similarity spaces: geometric models to predict people's similarity judgments with regard to whichever respect the given space is

[1]Different people will in general give somewhat different similarity judgments. For that reason, one might want to construct a similarity space for each person individually. However, that is often impractical, and non-individualized spaces—spaces typically constructed on the basis of similarity judgments coming from different people, and not pretending to represent anyone's personal similarity space—tend to work well enough for many purposes; see Douven (2016).

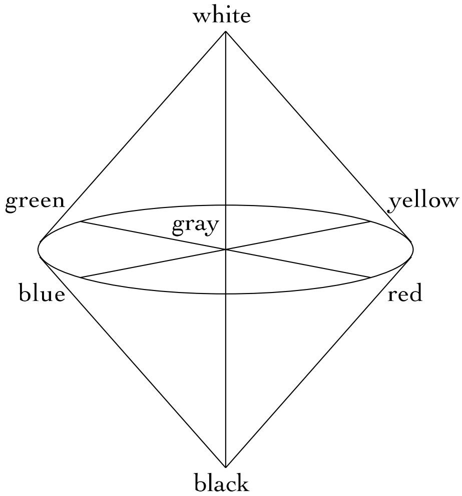


Fig. 1 An approximate representation of color space

meant to capture. But most of these spaces can be, and have been, further developed into *conceptual* spaces. That requires designating particular regions in a similarity space as representing specific concepts. So, for instance, SWEET is a specific region in taste space, and VASE is a specific region in a particular shape space. Concepts, in this view, are geometric objects in similarity spaces.

To make this more concrete, consider the representation of color space given in Fig. 1. Color space is three-dimensional, with hue being represented by the color circle that one sees in the middle of the space (this goes through all the colors in the rainbow), lightness being represented by the vertical axis (which goes from white to black through all shades of gray), and saturation (the "depth" of a color) being represented by distance from the vertical axis.[2] The labels in the figure are not to be interpreted as indicating the locations where the named colors are to be found in the space. Rather, they give an indication of roughly where the corresponding color *concepts* are to be found, which are regions—*sets* of points—in this space, not single points.

If concepts are regions in a similarity space, is every region in a similarity space a concept? Given that "concept" is a somewhat technical notion, we are probably free to answer the question in the positive. But what this question means to ask, of course, is whether every such region corresponds to a concept that might be of practical relevance to us, one that we would have real use for and that we would want to have a predicate for in our language. Consider that, mathematically speaking, any connected set of points in color space is a region in that space. And surely the vast majority of regions—"almost all," in the technical sense of measure theory—are not

[2] "Distance" in color space is usually taken to mean "Euclidean distance." Not all conceptual spaces are Euclidean spaces, but the issue of which metric is appropriate for which space need not detain us here. For details, see Gärdenfors (2000, Chap. 3) or Douven (2016).

worth having a name for or otherwise singling out; the vast majority strike us as corresponding to very unnatural and gerrymandered concepts at best.

This raises the further question of what distinguishes those concepts that pre-theoretically strike us as natural from those that do not do so. In Gärdenfors (2000, 71), we find the following tentative proposal:

CRITERION P

A *natural concept* is a convex region of a conceptual space.

To say that a region is convex is to say that, for any two points in the region, the line segment connecting those points lies in its entirety in the region as well. Gärdenfors (p. 70) defends CRITERION P as "a principle of *cognitive economy*; handling convex sets puts less strain on learning, on your memory, and on your processing capacities than working with arbitrarily shaped regions." He also notes that there is already a fair amount of evidence supporting the thought that the concepts we use do correspond to convex regions of similarity spaces.

However, none of this helps in answering the question of whether convexity is also a *sufficient* condition for naturalness. Gärdenfors (*ibid.*) is not entirely convinced that it is, and explicitly remarks that he "only view[s] the criterion as a *necessary* but perhaps not sufficient condition on a natural property." In fact, the word "perhaps" here understates the extent to which CRITERION P falls short of capturing our intuitive notion of naturalness vis-à-vis concepts.

To illustrate the problem, consider again the color space shown in Fig. 1 and note that any plane intersecting that space divides it into two regions, both of which are convex. Moreover, it can be shown that whenever the intersection of several convex sets is taken, the result is itself convex (Douven et al. 2013, 147). Hence, carving up color space by means of any random pick of planes intersecting that space will result in a partitioning of the space with only convex cells. This is so even though some of these cells could contain shades that any normal observer would deem as falling under different color concepts. If convexity were sufficient for naturalness, we would nevertheless be forced to say that each of those cells corresponds to a natural concept.

We leave the question of the characterization of naturalness open for now, but will return to it in Sect. 4. There it will be argued that, while CRITERION P on its own is not capable of defining naturalness, there are other principles very much like it, both in spirit and in motivation, such that together these principles stand a good chance of doing the job. First, though, we turn to an account of vagueness that was built on the conceptual spaces framework as presented in this section.[3]

[3] Note that, to date, it is still largely unknown how broadly applicable the conceptual spaces framework really is. It has been successfully used to model perceptual concepts and to some extent also more abstract concepts (e.g., Gärdenfors and Zenker 2013). Whatever the framework's limitations may be, it is clear that the current proposal, by relying so heavily on this framework, will face the same limitations.

3 Conceptual Spaces and Vagueness

If concepts are to be conceived as regions in similarity spaces, then how are we to accommodate the fact that many concepts are vague? If RED is a region in color space, this suggests that, for any shade, we can say that it is either red (it is represented by a point lying in the region representing RED) or not red. But that runs counter to everyday experience: there are shades that we cannot classify so well, that strike us as being neither quite red nor quite some other color. It would seem that we need something like *blurry* regions to represent color concepts, or other vague concepts. However, mathematically it is not clear what blurry regions might amount to.

Douven et al. (2013) present a formal account of vagueness cast within the conceptual spaces approach. The account actually builds on more than the framework of conceptual spaces by combining Gärdenfors' (2000) proposed combination of that framework with prototype theory (Rosch 1973) and the mathematical technique of Voronoi tessellations.

Not all readers will be familiar with Voronoi tessellations, but the basic idea behind them is very simple. Let S be a space and $P = \{p_1, \ldots, p_n\}$ a set of points in S. Then the Voronoi tessellation of S generated by P is the set of cells $\{c_1, \ldots, c_n\}$ with each c_i containing all and only points in S that are at least as close to p_i as they are to p_j, for all $j \neq i$. (For an example, see Fig. 2, left panel.)

Suppose we are given locations in color space of the prototypical colors. Then we can let those locations act as points generating a Voronoi tessellation of color space, which would gather the red shades around the RED prototype, the green shades around the GREEN prototype, and so on. An interesting observation Gärdenfors makes is that

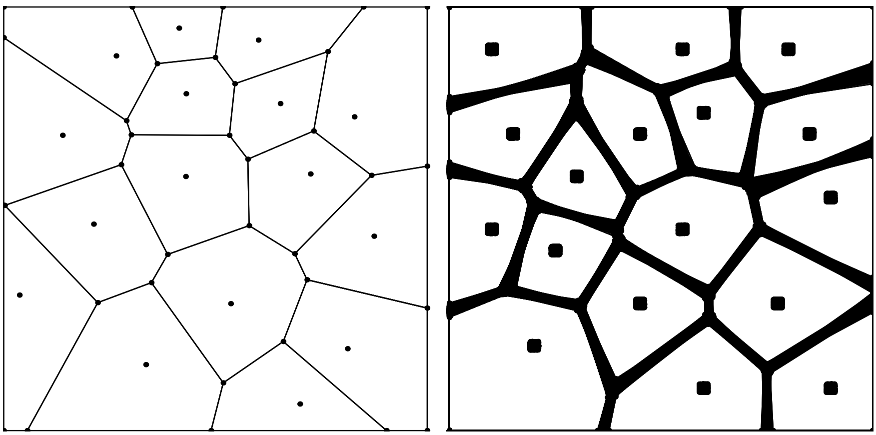


Fig. 2 Simple Voronoi tessellation (left) and collated Voronoi tessellation (right) on a two-dimensional space

this would automatically yield convex concepts, given that, necessarily, any cell of a Voronoi tessellation is convex (Okabe et al. 2000, 58).[4]

It might be thought that this answers the previous question regarding what makes a concept natural, the putative answer being that it is not convex concepts per se that are natural, but convex concepts resulting from a Voronoi tessellation generated by the relevant prototypes. But this cannot be right. After all, prototypes are not supposed to be prior to concepts. Rather, a prototype is said to be the best representative, the most typical instance, of a concept (Rosch 1973, 330; see also Rosch 1975). Thus, the concept must be there before the prototype can come into existence.

At first, it might also seem that Gärdenfors' add-ons to the conceptual spaces framework answer, by themselves, the earlier question about vagueness. Looking again at the left panel of Fig. 2, we see that on Gärdenfors' account concepts clearly have borderline cases: these are what constitute the borderlines between the regions, the points that are equally far from the prototypes of at least two different concepts. However, proposing this as an answer to the question of where vagueness comes from would lead to what Douven et al. (2013) have termed "the thickness problem." Open on your computer any drawing program that you are familiar with and go to the color menu. Then move the sliders (or dials, or whatever exactly the interface uses) to set the color coordinates in whichever color space the program uses (probably RGB space[5]) so that the resulting color strikes you as being borderline red/orange. Then, *very* probably, moving any of the sliders just a tiny bit in either direction will result in a color that *still* strikes you as being borderline red/orange. That would not be possible, however, if we were to build our account of vagueness directly on Gärdenfors' proposal. For then "almost all" (again in the measure-theoretic sense of that expression) small adjustments of the sliders would have to yield either a clear case of red or a clear case of orange. Put graphically, borderline cases tend to be surrounded by borderline cases, which would seem to require border *regions* instead of border*lines*.

Douven et al. (2013) have come up with an arguably non-ad hoc way of equipping conceptual spaces with such border regions. The non-ad hocness derives from the fact that their proposal starts from an independently plausible observation and then extends the technique of Voronoi tessellations in a natural way immediately motivated by that observation.

The observation is that, for many concepts, it is at best a convenient fiction to suppose that it has exactly one prototype, one best representative. Open your drawing program again, and adjust the color coordinate sliders until the color you see on the canvas strikes you as typically red. Now ever so slightly move one of the sliders. Whichever slider you move, and whether you move it to the left or to the right, chances are that the color on the canvas will still strike you as typically red. What

[4]This result holds for Euclidean spaces, such as color space and many other similarity spaces, but not for all spaces. Again, these details are left aside here.

[5]To forestall misunderstanding, it is to be noted that RGB space and other color spaces frequently used in applications software are not *similarity* spaces; they are not intended to represent human similarity judgments concerning color, and they are also as a matter of fact known *not* to represent such judgments.

goes for red goes for any other color, and indeed, *mutatis mutandis*, for most or even all other perceptual concepts. And it is not, or not necessarily, a matter of discriminability. I have no difficulty telling apart the taste of vanilla ice cream and that of strawberry ice cream, and yet both strike me as typical for the concept SWEET. (I am talking here of our local ice cream; where you live, the strawberries may not be as sweet.)

It thus appears that, for many conceptual spaces, we are better off thinking of them as having prototypical *regions* rather than prototypical *points*. If this is accepted, however, then what remains of the other part of Gärdenfors' proposal, the technique of Voronoi tessellations? Standard Voronoi tessellations have so-called generator *points*, not generator *regions*.

Douven et al. (2013) argue that there is still a use for Voronoi tessellations, provided that we slightly change the technique. In a nutshell, the modification they propose is this: Consider the set V of all possible selections of one single point from each prototypical region in a space and note that every member of that set generates a Voronoi tessellation on the space. Then, instead of looking at those tessellations one at a time, imagine how they are all simultaneously present in the space. Call the resulting construct "the collated Voronoi tessellation generated by V" on the given space. Douven et al. (2013) show that this construct still partitions the space into convex regions. For the purposes of explaining vagueness in the conceptual spaces framework, this constitutes an important step forward in that collated Voronoi tessellations have "thick" boundary regions, meaning that "almost all" borderline cases are surrounded by other borderline cases. For an illustration, see the right panel of Fig. 2.

Even if this picture yields a phenomenologically more plausible account of borderline cases and therefore of vagueness, the fact that, as the collated Voronoi tessellation in Fig. 2 shows, borderline cases are still sharply delineated from non-borderline cases (i.e., clear cases) could give cause for concern, simply because it seems unrealistic. We could again use a drawing program to verify that we do not tend to experience sudden transitions from, say, borderline cases of red to clear cases of red: where the clear cases end and the borderline cases begin is, for most or all vague concepts, itself a vague matter. In other words, the new proposal still seems to face the so-called problem of higher-order vagueness.

This problem can actually be dealt with in more than one way. Douven et al. (2013) account for it by reference to the imprecision of psychological metrics. But a more satisfactory solution to the problem is given in Decock and Douven (2014), which develops a measure of graded membership, and which explains our experience of higher-order vagueness by showing that there are no abrupt transitions from clear cases to borderline cases, given that the borderline cases neighboring the clear cases fall under the given concept to a degree still very close to 1 (and thus very close to the degree of any of the clear cases; see also Douven and Decock 2017).

To end this section, I mention that Douven et al. (2013) do note that their proposal has clear empirical content, and that in recent years the proposal has been put to the test and so far has passed all tests with flying colors. See Douven (2016), Douven et al. (2017), and Verheyen and Égré (2018).

4 Naturalness and Design

It is all well and good to have an account of vagueness in the conceptual spaces framework, even one that already enjoys some empirical support, but there remains a cloud over the account as long as it has not been shown that the conceptual spaces framework itself is tenable. As we saw in Sect. 2, a question that looms large over the framework is whether it has the resources to differentiate between natural and non-natural concepts. If ultimately we had to admit that, given this framework, any convex region in a conceptual space, however bizarre and gerrymandered it may appear pre-theoretically, represents a concept as natural as, say, GREEN, we would want to reject the framework, and the account of vagueness described in the previous section would go out of the window with it.

Douven and Gärdenfors (2018) have recently addressed the question of how to characterize natural concepts within the conceptual spaces framework. To do so, they take their cue from work on design thinking in theoretical biology. Biologists have been arguing that many traits and features of organisms, and also many structures of cells or even structures of molecules within cells, are best understood from a design perspective, specifically as being the best solutions to engineering problems. Douven and Gärdenfors argue that this same design thinking also applies at the cognitive level and promises to answer the question of what natural concepts are. Notably, their proposal is that natural concepts are concepts represented by the cells of an *optimally designed* conceptual space. Most of their paper then consists of (i) spelling out the notion of optimal design as it pertains to conceptual spaces, and (ii) illustrating and supporting the proposal by discussing a number of recent results concerning categorization from the cognitive sciences.[6]

To accomplish (i), Douven and Gärdenfors state a number of (what they call) design criteria and constraints, where design *criteria* hold generally for conceptual structures, independently of the representational format we choose for concepts (so, in particular, independent of the conceptual spaces format), while design *constraints* pertain explicitly to one specific type of conceptual structure, to wit, conceptual spaces. The criteria and constraints are all supposed to pertain to conceptual structures for creatures that, like us, have memories with limited storage capacity; are not able to detect arbitrarily small differences between stimuli; and seek to thrive in a world in which commodities are scarce and competition for those commodities can be fierce.

[6] Douven and Gärdenfors' approach is somewhat akin to Anderson's (1990, Chap. 3) rational analysis of categorization, which tries to understand categorization as the outcome of a procedure aimed at maximizing predictive success and in particular also invokes optimality considerations to identify what Anderson calls "basic level categories." Apart from the fact that Anderson is not concerned with conceptual spaces, another important difference is that he takes the world to be carved up independently of any human mental activity. That objective structure is to be discovered by us, and that is where—on his account—optimality comes into play. In contrast, Douven and Gärdenfors are concerned with showing what the natural concepts are, not—in first instance—how we latch onto those concepts. It is in answering the first question that they appeal to optimality considerations. (Thanks to Daniel Lassiter for pressing me on this).

The criteria and constraints are meant to help endow such creatures with conceptual structures that increase their chances of long-term survival and of reproduction.

Douven and Gärdenfors list the following design criteria:

PARSIMONY

The conceptual structure should tax the system's memory as little as possible.

INFORMATIVENESS

The concepts should be informative, meaning that they should jointly offer good and roughly equal coverage of the domain of classification cases.

REPRESENTATION

The conceptual structure should be such that it allows the system to choose for each concept a prototype that is a good representative of all items falling under the concept.

CONTRAST

The conceptual structure should be such that prototypes of different concepts can be so chosen that they are easy to tell apart.

LEARNABILITY

The conceptual structure should be learnable, ideally from a small number of instances.

As mentioned, the creatures to whose conceptual structures these criteria are supposed to pertain have limited memory capacity, which explains the presence of PARSIMONY on the list.[7] On the other hand, conceptual structures should be useful; in particular, they should allow the creatures to make sufficiently fine-grained distinctions between whichever items they may encounter, as their success will depend on this capacity—which explains INFORMATIVENESS. REPRESENTATION and CONTRAST are both motivated by the creatures' limited discriminatory capacities in combination with the fact that the creatures will need to avoid making classification errors as much as possible. Finally, the creatures will have to be able to use the conceptual structures after a relatively short time period following birth, which is why LEARNABILITY is on the list.

As for design *constraints*, CRITERION P is among them, now taken to derive from REPRESENTATION (convex regions make it easy to pick representative prototypes: just pick their centers, or small regions surrounding and including their centers) and PARSIMONY (recall that Gärdenfors presented CRITERION P as a principle of cognitive economy). There is also the constraint (adapted from the machine-learning literature on clustering; e.g., Kaufman and Rousseeuw 1990) that a conceptual space should be such that items falling under any of its concepts are maximally similar to each other and maximally dissimilar from the items falling under any of the other concepts represented in the same space. Following Regier et al. (2007), Douven and Gärdenfors label this constraint WELL-FORMEDNESS.

To make the foregoing more concrete, and to show how the above principles may help to fix a furnishing of a similarity space with concepts, suppose that a group of engineers are to endow an artificial creature with a system of concepts. They are

[7] Marzen and DeDeo (2017) formally show how cost considerations in organisms motivate parsimonious representational systems, even though this tends to come at the cost of greater inaccuracy.

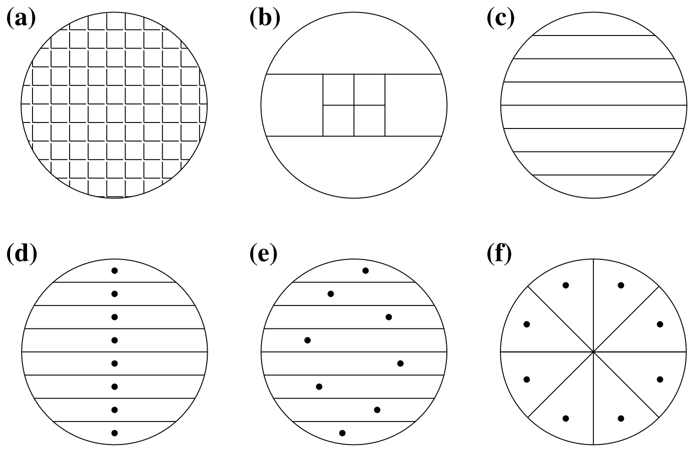


Fig. 3 Six candidate conceptual structures for a disk-shaped similarity space

asked to work under the assumption that the creature's limitations are relevantly similar to ours (the same limitations on storage capacity, discriminatory capacities, available time to familiarize itself with the conceptual system) and also that it has to operate under much the same pressures that we face, although at the point of design no specifics are known about the environment that it will inhabit. Thus the engineers are asked to solve a *constrained optimization* problem.

They tackle the task in a piecemeal fashion, starting with the subtask of providing the conceptual structure for one particular perceptual domain. By eliciting from the creature similarity judgments concerning the given domain, and by applying one of the statistical dimensionality-reduction techniques briefly mentioned in Sect. 2, they are able to ascertain that the corresponding similarity space is a perfectly round disk with a Euclidean metric defined on it.

In trying to determine the optimal conceptual structure for this space—optimal given the system's limitations—they pay heed to the design criteria and constraints stated above. It is clear to the engineers that these criteria and constraints will sometimes pull in different directions, so that their goal must be to find a conceptual structure that does best *on balance*.

The engineers begin by comparing the six conceptual structures in Fig. 3. Structure A does very well on INFORMATIVENESS: it endows the space with many concepts, and all parts of the space are covered equally. However, it is more than likely that A will overtax the system's memory and that it will therefore do poorly on PARSIMONY. Structure B does much better in the latter respect; but although the engineers believe that eight concepts may be enough for this similarity space, the coverage offered by B is very unequal: it allows the system to make relatively fine distinctions in the central part of the space but only coarse distinctions in the peripheral parts. Structure C, then, might be an acceptable compromise between INFORMATIVENESS and PARSIMONY.

The engineers realize, however, that structure C is still far from ideal, given that it makes it hard to jointly satisfy REPRESENTATION and CONTRAST. Consider D, which equips each concept in C with a prototype. D locates the prototypes in a perfectly symmetrical fashion, which seems to be the best way to make them representative—it minimizes the average Euclidean distance to all points in the concept corresponding to the given prototype—even though (especially for the concepts represented by the segments in the middle of the space) the prototypes are quite distant from the points that fall under the concept but lie close to the border of the space.

More problematic about D is that prototypes of adjacent concepts lie very close together, and lie much closer to each other than, for instance, to points in the same concept that are close to the border of the space. In fact, the prototypes lying nearest to the center of the disk lie closer to almost all other prototypes than to some points in the concepts of which they are the prototypes. So, the prototypes are not contrastive at all.

One could try to mend this defect by slightly shifting each of the prototypes, perhaps as done in E. But while this makes the prototypes a bit more distinct from each other, doing so comes at the expense of representativeness. Consider, for example, that the prototypes of the two middle-segment concepts are closer, and thus more similar, to almost *any* point in all the other concepts than they are to some of the points that lie in their "own" concept.

In short, while structure C may be an acceptable compromise between INFORMATIVENESS and PARSIMONY, it scores poorly on the counts of REPRESENTATION and CONTRAST: it does not allow the system to choose prototypes that are both sufficiently representative of their concepts while also being sufficiently different from the prototypes of other concepts.

Now consider structure F, which divides the space into eight concepts, like B and C do, and also offers equal coverage, providing another good compromise between INFORMATIVENESS and PARSIMONY. But F also offers room for making the prototypes both representative of their concept and distinct from each other. Of course, the prototypes can still be placed a bit more toward the center of the disk, which might make them more representative, or placed a bit closer to the border of the space, which would make them a bit more distinct. But that can all be left to the artificial creature itself. The engineers are to make sure that the creature can make such choices in the first place.

Surely, however, there are partitions of the disk-shaped similarity space other than F that have these virtues. For instance, G in Fig. 4 may appear to do quite well, too, in terms of PARSIMONY, INFORMATIVENESS, REPRESENTATION, and CONTRAST. But note that if CRITERION P is entailed by PARSIMONY and REPRESENTATION, as Douven and Gärdenfors (2018) argue, then G is ruled out, given that some of the concepts it creates are non-convex.

If we want to go along with Douven and Gärdenfors and claim that, starting from some obvious design principles, we may be able to identify the "natural" conceptual architecture for the disk-shaped similarity space, then structure H shown in Fig. 4 seems to pose more of a challenge. After all, while F seems to satisfy the engineers' design principles, so does any rotation in the space of the conceptual structure *cum*

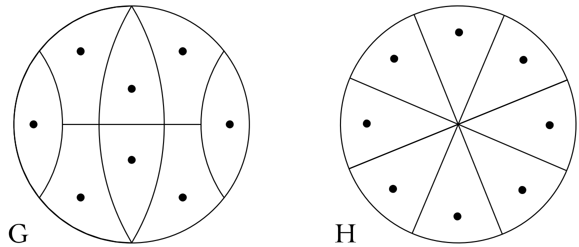


Fig. 4 Conceptual structure with some nonconvex concepts (G) and rotation of structure F from Fig. 3 (H)

prototypes—such as H. We thus seem compelled to admit that there is an abundance of natural structures, and so an abundance of natural concepts (even if not every convex region corresponds to a natural concept), and that seems just wrong: natural concepts are supposed to be *sparse* (Lewis 1983).

The response to this challenge is two-pronged: First, I claim that the problem of rotational variants stems from the fact that we are considering a fictional, abstract similarity space. It might immediately be countered that this cannot be right, for look again at perceptual color space, as shown in Fig. 1, which is not a fictional, abstract similarity space. It seems that design principles of the kind considered so far are fundamentally unable to determine a conceptualization of that space yielding natural concepts, given that any rotation along the luminance axis is structure-preserving in the relevant sense: any non-trivial rotation along that axis of a structure satisfying the design principles introduced so far will satisfy those principles too, although it will yield different color concepts. It is to be stressed, however, that Fig. 1 only shows an *approximation* of perceptual color space. The approximation is quite good, yet real color space is not nearly as symmetric as the spindle shown in Fig. 1. Just google images of CIELAB space and you will see what I mean.[8] Important work by Regier et al. (2007) suggests that the asymmetry of CIELAB space in conjunction with WELL-FORMEDNESS may even be enough to determine a *unique*, or at least *near-unique*, partitioning of that space into color concepts; see also Douven (2017) and Jraissati and Douven (2017).

Second, the above example is meant to illustrate how, in the context of the conceptual spaces framework, design considerations may be able to pin down a conceptual structure for a similarity space, once that similarity space has been determined. In other words, the point is to show how (taking into account the constraints under which a creature is to operate) there may be an optimal architecture for each of the creature's similarity spaces, where this architecture can then be said to yield a set of concepts that are natural. Douven and Gärdenfors' claim is *not* that the five design criteria they introduce are guaranteed in every case to be jointly sufficient for the task of identifying the optimal structure. In fact, they leave open the possibility that

[8] Or google images of CIELUV space, which is another, slightly different, perceptual color space; see Malacara (2002, pp. 86–90) for details and for a discussion of how CIELAB and CIELUV spaces differ.

the design criteria that determine optimality cannot be found strictly through a priori reflection.

As said, Douven and Gärdenfors (2018) paper consists of two main parts: one in which they propose, and argue for, the design criteria and constraints mentioned above, and another in which they muster experimental findings reported in other publications—typically also in very different contexts—which actually yield support for the design view of naturalness. I have already mentioned Regier et al. (2007), which shows how WELL-FORMEDNESS helps explain the commonalities we find in color lexicons from cultures across the globe, as registered in the World Color Survey (see Cook et al. 2005). Kemp and Regier (2012), Xu and Regier (2014), and Xu et al. (2016) present similar results concerning kinship categories, numerical systems, and container categories, respectively. Finally, there is work on social learning by Jäger (2007), van Rooij and Jäger (2007), and Xu et al. (2013), which is discussed at some length in Douven and Gärdenfors (2018) and which provides evidence for holding that INFORMATIVENESS, PARSIMONY, and LEARNABILITY are operative in the cognitive domain.

5 Rationality and Vagueness

According to Douven and Gärdenfors (2018), natural concepts are those represented by the cells of an optimal partitioning of a similarity space. In other words, the natural partitioning into concepts of a given domain is the one that is most rational, in that it is the one that a good engineer would provide if she were asked to design our conceptual apparatus for us. So, on this account, there is a clear link between rationality and categorization. On the proposal to be made now, however, that account only describes the first step toward categorization. It is as if an architect had come up with a first rough sketch of your to-be-built house, reflecting a rational division of the available space, based on your needs and desires, but leaving out some important finer details. The filling-in of those details is itself subject to rationality principles.

I am actually not sure whether "details" is the right word in the present case. Here, the details concern vagueness and prototypes, which are of central importance to categorization. Douven and Gärdenfors (2018) explicitly bracket the former issue. And while, on their account, there is a connection between optimal categorization and prototypes—most notably, via REPRESENTATION and CONTRAST—prototypes themselves are not yet in the picture; for Douven and Gärdenfors' purposes, exactly where the prototypes end up being located in a space is not relevant to what makes concepts natural.

In this section, I consider what a straightforward extension of Douven and Gärdenfors' account of naturalness might look like, where the extension should be sensitive both to the fact that most concepts are vague and to the fact that conceptual spaces are equipped with prototypes. The claim is not just that the account of vagueness from Sect. 3 fits naturally with the account of naturalness from Sect. 4, but that their combination flows from rationality considerations and that, consequently, vagueness can be thought of as resulting from those considerations.

Specifically, I argue that (i) it is rational to equip conceptual spaces with prototypical regions rather than with just prototypical points; (ii) it is rational to put each prototypical region at a location such that the prototypes are as representative of their concepts as possible and as different as possible from the prototypes of the other concepts in the space; and (iii) it is rational to obtain the final carving-up of a similarity space by constructing a collated Voronoi tessellation, in the way explained in Sect. 3, on the basis of the prototypical regions placed in line with (ii). Clauses (i) and (ii) together form the second step of categorization, and clause (iii) is the third and final step of categorization.

The Second Step of Categorization. In Douven and Gärdenfors (2018) proposal, REPRESENTATION and CONTRAST refer to prototypes, which throughout their paper are supposed to appear as isolated points in a conceptual space. We know from Sect. 3 that this can only be an idealization, and that it is more realistic to think of conceptual spaces as having prototypical regions. The latter claim is justified by a wealth of evidence, both from introspection and from experimental studies, showing that going by people's judgments, many concepts have multiple "best" instances. Importantly, that many concepts have prototypical regions (as opposed to prototypical points) is not *just* true as a matter of empirical fact; it is a fact for which there exists a rational explanation.

In the psychological literature it is often said that we construct prototypes by abstracting from concrete instances we have encountered, as if we create a summary of those instances (see Gärdenfors 2000, Chap. 4.5). If so, learning a concept C is strictly a matter of being shown certain items, which are labeled as C by one's parents or teachers, whence one should try to figure out what makes those items relevantly similar.

No doubt this is part of the process of concept acquisition. But it is far from the whole story. Most notably, it completely ignores the educational role that the term "typical" and others play in the practice of teaching the meaning of a word. We *often* use "typical" when we want to convey the meaning of a word to a child: we do not just tell her that that thing over there is yellow; we tell her that the thing is *typically* yellow, or that this lemon's taste is *typically* sour, and so on. In other words, often we help a child to acquire a concept not just by showing it various items falling under that concept, labeling them with whatever predicate we use for the concept, and hoping that the child will somehow "get" what a best representative of the concept looks like; but by directly presenting some best representative to it, stressing that this *is* a best representative of the relevant concept. That practice will greatly help to speed up the learning process.

Naturally, multiple prototypes for a concept will offer more opportunities for pointing out what is typical for that concept. And the more such opportunities we have, the more we will be able to speed up the learning process. In particular, we are then likely to teach the concept much faster than if there were just one prototype, with correspondingly fewer opportunities to use the word "typical" in relation to the concept, so that we have to rely much more on the child or student to discover for herself the important commonalities of the items we are grouping under the same label.

In fact, suppose there were exactly one BLUE prototype. In color space, this would correspond to a single point, which has measure 0. The chance that we would ever encounter exactly that shade, and thus be able to point it out to a child or student, would be essentially nil. By contrast, the prototypical BLUE region as identified in Douven et al. (2017) has a positive measure, and we can be as good as certain that we all have encountered plenty of shades of blue falling into that region.[9]

It is worth mentioning that, so far, all empirical evidence on prototypical regions is consistent with the assumption that such regions are convex (see, e.g., Douven 2016 and Douven et al. 2017). LEARNABILITY provides independent and a priori support for that assumption. After all, if we can assume that prototypical regions are convex, then the learning of which regions in a space represent the prototypes becomes very efficient: by learning of just a handful of shades that they are typical instances of, say, BLUE, we then automatically learn that every shade represented by a point inside the convex hull of those typical instances is a typical instance of BLUE as well.

The second step of categorization also concerns the question of where in a given space—already equipped with the partitioning resulting from the first step—to *place* the prototypical regions. Here, too, there are *rational* decisions to be made: we want the prototypical regions to be representative of their concepts, and we also want them to be sufficiently distinct from each other, to facilitate memorization and to limit the chance of misclassifications, as previously explained. It is a matter of rational design to ensure that prototypical regions *can* be located in a space so that they are both representative and contrastive, as REPRESENTATION and CONTRAST jointly imply, precisely because it is rational to *have* prototypical regions that are representative and contrastive.

As an aside, it is to be noted that our discussion of structure F above made it clear that rationality may leave some leeway here. It can happen that, by making the prototypical regions less central, and so less representative, they become more distinct from one another. And if such a trade-off is to be made, there may be no single best way to do it. Possibly, different rational people will make different trade-offs in such cases, which would also explain why we find some individual variation in people's responses to questions asking them to designate the most typical item or items falling under a given concept (see Berlin and Kay 1969; Douven et al. 2017).

The Third Step of Categorization. There is a third, and final, step of categorization in my proposal here, namely, that of generating a collated Voronoi tessellation from the prototypical regions, as described in Sect. 3. This procedure remains unaltered here. It is worth pointing out, however, that the procedure has a rational basis as well.

Defending his proposal of combining the conceptual spaces framework with prototype theory and the technique of Voronoi tessellations, Gärdenfors (2000, 89) convincingly argues that it yields a "cognitively economical way of representing

[9] To be entirely precise, Douven et al. (2017) give a number of different estimates of the prototypical BLUE region in CIELUV space, but what is said above holds given any of those estimates.

information about concepts," the reason being that it is much more cost-effective to keep in memory the locations of the prototypes and to compute the categorization of any given point in the space by finding the prototype nearest to it than to remember the category of each point in a space.

Since we can just as easily measure distances between points as those between regions, or between regions and points, Gärdenfors' motivation for appealing to the technique of Voronoi tessellations is also valid for the extended technique of collated Voronoi tessellations. Admittedly, as can easily be shown, the procedure of just looking which prototypical region a given point is nearest to will not always tell you whether that point is a clear case or a borderline case. For that, one should ask whether the point is closer to *every* point in that region than it is to *any* point in *any* other prototypical region in the same space. But answering *that* question is probably *not* very cost-effective. So, when dealing with prototypical regions, the designated procedure can at best serve as a heuristic.

However, the procedure should still be good enough for practical purposes. And in any event, we may not be able to improve upon it, given that (as briefly mentioned in Sect. 3) psychological measures tend to be imprecise. So, while following the heuristic may not always give an exact answer as to which concept a given point falls under, trying to answer this question in the more involved way will not always yield a precise answer either—because of the inherent inexactitude of measuring distances in whichever space the point lies—and may in general not even yield a more exact answer than the one we get from the heuristic.

Discussion and Directions for Future Work. In short, the proposal is that categorization can be thought of as consisting of three steps: in the first step, design criteria and constraints—which are at bottom rationality principles—determine a provisional partition of a given space; in the second step, prototypical regions are put into place, guided by the desiderata of representativeness and contrastiveness, where the fact that we are referring to regions has a rational motivation itself; and in the third step, the final conceptual architecture is achieved by generating a collated Voronoi tessellation from the prototypical regions, a procedure that also has a rational backing, as explained above. The process of going through these steps is illustrated in Fig. 5 for an abstract two-dimensional similarity space. (To forestall misunderstanding, there is in general no reason to expect all prototypical regions to have the same size and shape—as the second panel of Fig. 5 might seem to suggest—nor that the boundaries regions obtained in the final step will always be "well-aligned" with the borderlines obtained in the first step, as a comparison of the left and right panels might suggest.)

It cannot be stressed enough that what I have tried to accomplish in this paper is *not* to present a plausible phenomenology of the process of categorization. Rather, I have tried to give content to the notion of an optimal cognitive architecture. Douven and Gärdenfors (2018) have already argued that the notion of naturalness as it pertains to concepts is to be understood in terms of optimally partitioned similarity spaces. However, they admit that their proposal cannot be the complete story. In particular, vagueness is explicitly set aside, and the location of prototypes receives no discussion. The current proposal is meant to round out Douven and Gärdenfors' account of

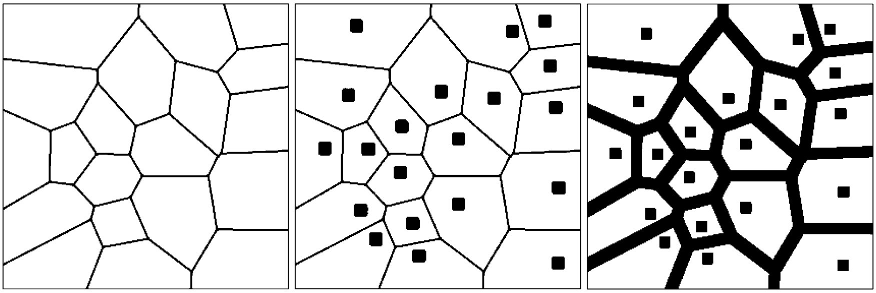


Fig. 5 Results of determining the rough conceptual architecture of a given similarity space (left), the locations of the prototypical regions in that architecture (middle), and the collated Voronoi tessellation generated by those regions (right)

naturalness, or at least to take it some important steps further. Natural concepts are those that reside in optimally designed conceptual structures, where it is now seen that the vagueness of concepts also results from the optimality—that is, *rationality*—of the design.

I want to remain neutral, at least for now, on how optimality of the kind at issue is achieved. It is clear that we do not consciously take design decisions to arrive at our concepts. For all that has been said, the verdict that *this* is a good spot for placing a prototype may well manifest itself "from the inside" as an overwhelming sense of typicality. But then again, much or even all of what leads us to have the concepts we have may be hard-wired into our physiology, or derive from mechanisms that are hard-wired. To repeat, the point I want to make is that we can make sense of the notion of an optimally designed conceptual architecture, and that we can do so in a way that makes vagueness a matter of rationality rather than some defect for which we may or may not be able to find an excuse. But the rationality involved is not necessarily *ours*. Rather—to come back to an image from the introduction—the point is that if somehow logicians and linguists had devised our language, they would, considering our physiology as well as the world we inhabit, probably have come up with a language that exhibits vagueness much in the way and to the extent that spoken languages actually do.

In closing, I would like to mention that while important parts of the proposal already enjoy empirical support—there is the experimental work on vagueness cited in Sect. 3, and Douven and Gärdenfors (2018) discuss at length evidence for their proposal concerning naturalness—so far little research has been done on the question of which principles may determine the locations of prototypes or prototypical regions in conceptual spaces. As mentioned, Douven and Gärdenfors leave open the possibility that the design criteria and constraints actually at work in structuring conceptual spaces are not exactly the ones they present in their paper, or are not restricted to those presented. Similarly, it is plausible that considerations of contrastiveness and representativeness play a role in determining where to locate prototypes in a space, but this is by no means certain.

One way to find out would be to see whether we can get good predictions of where the prototypical regions are to be found in a given space by assuming that their locations strike the best balance between (i) being as centrally as possible located in the respective concepts and (ii) being as distant from each other, on average, as is compatible with their still lying in their concepts. To give a concrete example, suppose we knew how the basic color terms partition CIELAB space and also where in that space the prototypical regions are located. That constellation of prototypical regions could then be compared, in terms of both representativeness and contrastiveness, with other possible constellations, and we might find that on balance none of those other constellations do as well as the actual one.

More concretely still, note that we could randomly relocate each prototypical region within its concept, and that we could consider, in simulations, thousands or perhaps millions of such random relocations. If it then turned out that those relocations which result in higher scores on one count (representativeness or contrastiveness) tend to result in lower scores on the other count, that would offer strong support for the part of our proposal that so far is still largely unsubstantiated.

Contrary to what many might think, the partition of CIELAB space that this test would require is not yet known, nor is the constellation of the prototypical regions in that space.[10] This is mainly due to the fact that color-naming studies have hitherto been conducted only on the "hull" of that space, typically by using the 330 Munsell chips that were used in the World Color Survey—chips that are all at maximum saturation for their hue–value combination. But color-naming studies, as well as studies asking for typicality judgments that use stimuli sampled from throughout the space, are currently underway, and the data from those studies should enable us to perform the test just described, and should thereby allow us to verify or falsify, as the case may be, claims concerning the design principles behind the locations of prototypical regions.[11]

References

Anderson, J. R. (1990). *The adaptive character of thought*. New York: Psychology Press.

Berlin, B., & Kay, P. (1969). *Basic color terms*. Stanford, CA: CSLI Publications.

Borg, I., & Groenen, P. (2010). *Modern multidimensional scaling* (2nd ed.). New York: Springer.

Cook, R. S., Kay, P., & Regier, T. (2005). The world color survey database: istory and use. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 223–242). Amsterdam: Elsevier.

Decock, L., & Douven, I. (2014). What is graded membership? *Noûs, 48,* 653–682.

Douven, I. (2016). Vagueness, graded membership, and conceptual spaces. *Cognition, 151,* 80–95.

[10] Douven et al. (2017) present the empirically determined locations of the BLUE and GREEN prototypical regions in CIELUV space; the locations in that space of other prototypical regions are still unknown.

[11] I am greatly indebted to Richard Dietz, Daniel Lassiter, Christopher von Bülow, and an anonymous referee for very valuable comments on previous versions of this paper. I am also grateful to an audience at the Ruhr University Bochum for stimulating questions and remarks.

Douven, I. (2017). Clustering colors. *Cognitive Systems Research, 45,* 70–81.
Douven, I., & Decock, L. (2017). What verities may be. *Mind, 126,* 385–428.
Douven, I., & Gärdenfors, P. (2018). What are natural concepts? A design perspective. *Mind & Language* (in press).
Douven, I., Decock, L., Dietz, R., & Égré, P. (2013). Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic, 42,* 137–160.
Douven, I., Wenmackers, S., Jraissati, Y., & Decock, L. (2017). Measuring graded membership: The case of color. *Cognitive Science, 41,* 686–722.
Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge MA: MIT Press.
Gärdenfors, P., & Zenker, F. (2013). Theory change as dimensional change: Conceptual spaces applied to the dynamics of empirical theories. *Synthese, 190,* 1039–1058.
Jäger, G. (2007). The evolution of convex categories. *Linguistics and Philosophy, 30,* 551–564.
Jraissati, Y., & Douven, I. (2017). Does optimal partitioning of color space account for universal color categorization? *PLoS One, 12*(6), e0178083. https://doi.org/10.1371/journal.pone.0178083.
Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data*. Hoboken, NJ: Wiley.
Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science, 336,* 1049–1054.
Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 85,* 445–463.
Lewis, D. K. (1983). New work for a theory of universals. *Australasian Journal of Philosophy, 61,* 343–377.
Malacara, D. (2002). *Color vision and colorimetry: Theory and applications*. Bellingham, WA: SPIE Press.
Marzen, S. E., & DeDeo, S. (2017). The Evolution of Lossy Compression. *Journal of the Royal Society, Interface, 14,* 20170166. https://doi.org/10.1098/rsif.2017.0166.
Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations* (2nd ed.). New York: Wiley.
Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences USA, 104,* 1436–1441.
Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4,* 328–350.
Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532–547.
Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology, 69,* 1996–2019.
van Rooij, R., & Jäger, G. (2007). Language structure: Psychological and social constraints. *Synthese, 159,* 99–130.
Verheyen, S., & Égré, P. (2018). Typicality and graded membership in dimensional adjectives. *Cognitive Science* (in press).
Xu, Y., & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
Xu, J., Dowman, M., & Griffiths, T. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B, 280,* 20123073.
Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science, 40,* 2081–2094.

Semantic Indecision

Timothy W. Grinsell

Abstract Linguistic vagueness results from aggregating many judgments into one. Thus, vagueness is a type of decision problem, and the sorites paradox—most famous of vagueness phenomena—is a paradox of collective decision making. The sorites paradox reflects this judgement-aggregating character of vagueness. In the case of vague predicates like *healthy*, the "judgments" are the rankings of contextually salient entities along some scale (like "health") and the "dimensions" are relevant criteria (like blood pressure and heart rate). The sorites paradox is paradoxical because it tracks changes along one dimension ("a person with slightly higher *blood pressure* than a healthy person is healthy") while ignoring others. The aggregation of many judgments into one is also a feature of Condorcet's paradox, a paradox of collective decision making. If three equal blocks of voters have certain preferences among the available candidates, then the judgment-aggregating procedure—a vote—fails to deliver an outcome. Condorcet's paradox and the sorites paradox are two sides of the same coin. Both paradoxes arise from the decision problems inherent in aggregating many judgments along many dimensions. And in both cases, plausible constraints on the aggregation process—well developed in the theory of social choice—lead to paradox.

David Lewis famously argued that "vagueness is semantic indecision," (Lewis 1986, p. 213).

> The reason it's vague where the outback begins is not that there's this thing, the outback, with imprecise borders; rather there are many things, with different borders, and nobody has been fool enough to try to enforce a choice of one of them as the official referent of the word 'outback.'[1]

Semantically, vague predicates like *outback* may be understood as choice functions of the sort addressed by rational choice theory, in that they choose among a set

[1]Lewis articulated this view in support of the argument that "[t]he only intelligible account of vagueness locates it in our thought and language" rather than in the world.

T. W. Grinsell (✉)
Holwell Shuster and Goldberg LLP, New York, NY, USA
e-mail: twg@uchicago.edu

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*,
Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_8

of entities according to consistent criteria (Mas-Colell et al. 1995). In these types of decisions, it is sometimes impossible to choose—a paradox of choice. That is, vagueness is a decision problem, and vagueness effects result from paradoxes of choice.

The vagueness effects considered here include participation in the sorites paradox (as illustrated in (1)), borderline cases (which is the biggest little city, or the shortest tall man?), and higher-order vagueness (related to the issue of borderline cases: when is someone *definitely tall* as opposed to just *tall*?[2]). These phenomena reflect the difficulty of drawing a boundary between the extension and the anti-extension of a vague predicate—between, for example, the *tall* things and the non-*tall* things.

For example, Reno, Nevada in the United States claims to be "the biggest little city in the world." Consider the following line of reasoning:

(1) Premise 1. A city with a population of 1 is a little city.
Premise 2. A city with a population of 1 more than the population of a little city is a little city.
Conclusion. A city with a population of 8,000,000 is a little city.

Iterating the reasoning of the second premise suggests that if Reno is a little city, then so (eventually) is New York.

The reasoning in (1) has gone wrong in at least two ways. First, the indifference relation implied by the second premise ($a \sim_{\text{little}} b$ in (2) below) is intransitive. The second premise reflects that vague predicates are "tolerant." Kamp (1981) explains tolerance like this: suppose the objects a and b are indistinguishable in the respects relevant to some property P; then either both a and b satisfy P or neither of them does.[3]

(2) Premise 2: $\text{little}(a) \wedge a \sim_{\text{little}} b \rightarrow \text{little}(b)$

A vague predicate like *little* is tolerant in this sense. However, the truth of both the first premise and the conclusion tells us that the relation "indifferent with respect to size," represented by $\sim$, is not transitive. Reno may be observationally indistinguishable from Cincinnati with respect to population, and Cincinnati from Honolulu, but Reno and Honolulu may noticeably differ in size.

A second and related way in which (1) goes wrong is that small changes along one dimension of size (like population) do not translate smoothly into changes in the meaning of the predicate *little*. Relative to cities, at least, *little* may be sensitive to

[2] As Williamson (1994) shows, the presence of borderline cases is not sufficient to ensure vagueness. For example, it may be possible to define a predicate *definitely-tall** such that people above 180 cm in height fall into its extension, people 170 cm or below do not, and we remain noncommittal about the rest. The people between 170 and 180 cm in height might fairly be called borderline cases, yet *definitely-tall** is not vague in the same way as its natural-language counterpart *definitely tall*. In particular, while a predicate like *definitely-tall** has an identifiable boundary between the definitely-tall* entities and the borderline cases—180 cm—the natural-language *definitely tall* does not. This is the essence of higher-order vagueness: boundarylessness in a predicate's extension.

[3] Kamp in fact frames tolerance in terms of "observational" indistinguishability. This point appears to relativize indistinguishability to our goals or interests (Fara 2002). I return to this point below.

changes in population, geographic area, or attitudes regarding the norms prevalent in little cities (i.e., friendliness, etc.). Small changes along the single dimension of population do not appear to change the value of ⟦little(Reno)⟧ from true to false. Rather, the value of ⟦little(Reno)⟧ appears to change with sufficiently large changes in population. The mapping from one dimension of *little* to the meaning of *little* appears to be discontinuous.

Both of these properties—intransitivity and discontinuity—are problems that also appear in circumstances of collective choice. This chapter applies social choice theory, the branch of choice theory concerning collective decision-making, to account for some cases of linguistic vagueness. The Marquis de Condorcet observed that “cycles” (or intransitivities, e.g. A is preferred to B is preferred to C is preferred to A) may arise in a collective body like a legislature whenever such a body tries to choose among three or more alternatives. In 1950, Kenneth Arrow proved that (under certain conservative assumptions) there is *no* collective decision-making procedure that is guaranteed to avoid cycling. Related generalizations of Condorcet’s work have also shown that collective decisions tend to be discontinuous: if a voter changes her preferences in a minor way, the output of the collective decision may change in a major way.

Collective choice therefore provides a framework to analyze phenomena displayed by vague prediactes. Evidence for the social choice view comes from properties of “multidimensional” adjectives like *healthy*. These adjectives involve evaluations of multiple different criteria, such as blood pressure or cholesterol in the case of *healthy* used in an appropriate context (this is why it is possible to say things like *healthy with respect to cholesterol*). Speakers compare contextually relevant entities according to these criteria, counting as *healthy* the entities that rank sufficiently high on sufficiently many, sufficiently important dimensions.

Understood this way, these adjectives are collective choice functions, subject to the same limitations constraining Condorcet’s legislature. Replace Arrow’s voters with different dimensions and Arrow’s candidates with contextually relevant entities, and Arrow’s impossibility result reconstructs itself in the semantics of gradable adjectives like *healthy*.

If this is right, then the intransitivity represented by the sorites paradox and the intransitivity represented by cycling arise from the same source, the aggregation of many judgments into one. The cycling problem is a failure of transitivity: if the legislature prefers proposal A to B, B to C, and C to A, the legislature’s preferences are intransitive. Similarly, the sorites paradox is widely interpreted to demonstrate the intransitivity of the “indifference” relation. It is not true, for example, that if A and B are ranked indifferently with respect to health, and if B and C are ranked indifferently with respect to health, then A and C are ranked indifferently with respect to health. Other vagueness effects, like borderline cases and higher-order vagueness, likewise flow from the demands imposed by any reasonable collective decision-making function.

Vagueness may therefore be indecision. And this indecision is semantic in nature: vagueness effects follow from standard assumptions about the meaning of gradable adjectives plus reasonable assumptions about aggregating many scalar dimensions of meaning into one. There is no answer to the question "is Reno the biggest little city?" because this answer involves a decision made impossible by the semantics of the gradable adjective *little* itself.

The remainder of this chapter is structured as follows. In Sect. 1, I introduce social choice and its limitations. In Sect. 2, I develop a choice-functional semantics for gradable adjectives like *healthy*, and in Sect. 3 I show that these adjectives inherit the limitations on collective choice. In Sect. 4, I demonstrate how these limitations account for vagueness phenomena like the sorites paradox.

1 Limitations on Collective Choice

Cycles may arise in a collective body like a legislature whenever such a body tries to choose among three or more proposals. In (3), for example, pairwise majority vote delivers the election of Candidate A over Candidate B (Blocks 1 and 2 against Block 3) and Candidate B over Candidate C (Blocks 1 and 3 against Block 2). By similar reasoning, Candidate C beats Candidate A (Blocks 2 and 3 against Block 1). As a whole, the legislature's preferences cycle: it prefers A to B to C to A.

(3) Block 1: Candidate A $>$ Candidate B $>$ Candidate C
Block 2: Candidate C $>$ Candidate A $>$ Candidate B
Block 3: Candidate B $>$ Candidate C $>$ Candidate A
Election: Candidate A $>$ Candidate B $>$ Candidate C $>$ Candidate A

Kenneth Arrow showed that there is no collective decision procedure that respects certain reasonable assumptions and avoids this sort of intransitive behavior (Arrow 1950). Let $X = \{x, y, z, \ldots\}$ be candidates in a voting procedure and let $N = \{1, \ldots, n\}$ be the set of voters. By hypothesis, these voters have transitive preferences. Arrow's assumptions are set out informally in (4).

(4) a. Unanimity: If every voter prefers alternative x to y, the collective choice prefers x to y.
b. Independence of Irrelevant Alternatives: The collective choice between x and y depends only how the voters rank x and y.
c. Unrestricted Domain: The voters can rank the relevant candidates in any way.
d. Non-Dictatorship: No one voter determines the result of the collective choice.

(5) Arrow's Theorem: There is no complete and transitive collective decision procedure f that respects all four constraints.

If Arrow's constraints are obeyed (and they have some normative appeal in the voting context), the collective decision procedure will be incomplete or intransitive (or both) (5).

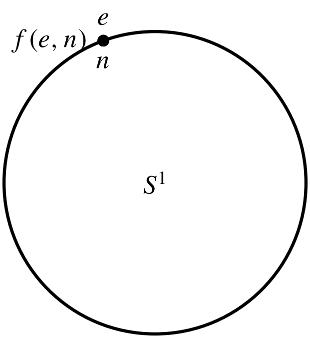
In addition to intransitivity, a collective decision procedure f may display discontinuities. That is, small changes in the voters' preferences do not necessarily result in a small change in the outcome (Gaertner 2009). For example, Gaertner discusses the following example: consider voters 1–5 with preferences among the alternatives v, x, y, z in descending order, as in (6).

(6)

voter 1	voter 2	voter 3	voter 4	voter 5
x	y	z	x	z
y	v	v	v	x
z	x	y	z	v
v	z	x	y	y

Alternative x is the "Condorcet winner" in this case because it beats all the other alternatives by majority vote in a pairwise contest. However, if the order of x and z are reversed in voter 2's preference ranking—voter 2's least prefered alternatives—then z is the Condorcet winner. A small change in the voters' preferences results in a large change in the electoral outcome.

This behavior was generalized by Chichilnisky (1982). Suppose two people, n and e, want to go camping along the shore of a perfectly circular lake.[4] They may prefer the same geographic location along the shore, or they may not. If they agree on the location, that is where they will camp. This is an analogue to Arrow's Unanimity.



(7)

And the decision will be anonymous, meaning that it will not depend on who chose which location. This is an analogue to Arrow's Non-dictatorship. In (8), the choices of n and e may be exchanged with no change in the resulting collective choice.

[4]This example is presented by Saari (1997).

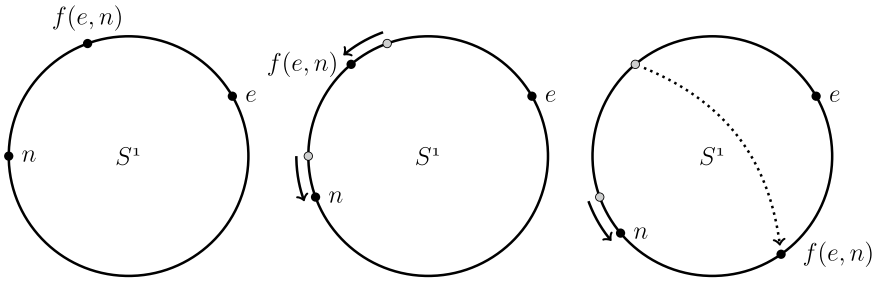
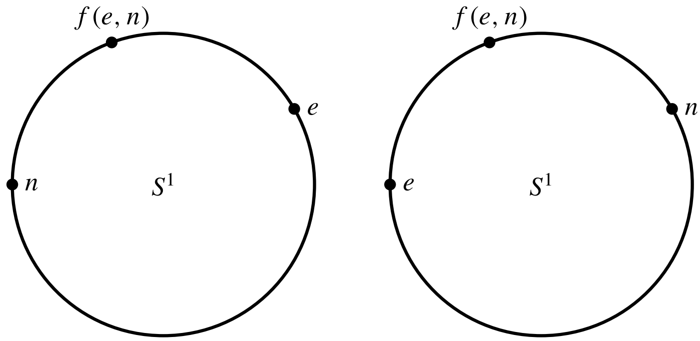
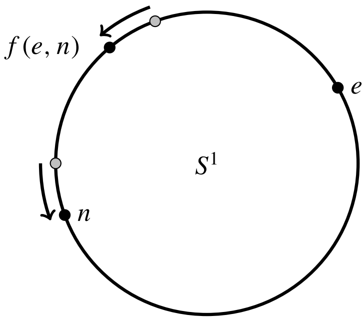


Fig. 1 Aggregation of camping location preferences along a circular lake (Grinsell 2014)



(8)

Finally, the decision rule should be "relatively insensitive to small changes in individual preference" (Chichilnisky 1982, p. 337). This last requirement is continuity.[5] In (9), a small change in n's preferred location results in a small change in the resulting collective choice.



(9)

Chichilnisky showed that there is no unanimous, anonymous, and continuous collective choice function f (10). To get a feel for why this should be, hold e fixed. As n moves continuously in a counterclockwise direction, so does $f(e, n)$, until n reaches the antipode of e. Then, $f(e, n)$ jumps to the other side of the lake (Fig. 1).

[5] Baryshnikov (1997) emphasizes the relationship between Independence of Irrelevant Alternatives and continuity.

(10) Chichilnisky's Theorem: There is no continuous aggregation rule $f : S^1 \times S^1 \to S^1$ that satisfies unanimity and anonymity.

As discussed in the next section, these two limitations on collective choice, intransitivity and discontinuity, are both reflected in vagueness phenomena.

2 Choice Theory for Adjectival Semantics

The problem of sorting the *healthy* entities from the non-*healthy* entities is equivalent to the problem of choosing the *healthy* entities from the rest. This view is similar to that presented by van Rooij (2011, p. 140). He defines a *context structure* $\langle I, C, V\rangle$ with I a nonempty set of individuals, C the set of finite subsets of I, and a valuation V that assigns to each $c \in C$ those individuals in c that count as being P, for some predicate P. Then, $P(c) = \{x \in c \mid x \in V(P, c)\}$. The idea, as van Rooij explains, is that "P is thought of as a choice function, selecting the *best* elements of c."

In addressing the problem of vagueness, the choice-functional approach is explicit about the "choice rule," how a set of entities (the domain of individuals) is sorted into P and its complement. For example, and as a first approximation, let the denotation of *happy* in some context c be as in (11).

(11) $[\![\text{happy}]\!]^c = \{x \mid \forall y \in c,\ xRy\}$

In order to account for vagueness, we want to know what properties the binary relation R has—what is the choice rule for happiness?

If the scale for the "positive" use of *happy* is the same as the scale for the comparative *happier*, then the choice rule R is already significantly constrained. For example, it should be possible to derive from R a transitive strict comparative relation $>$ for use in the semantics of comparatives like *Sam is happier than Sonia.* Along these lines, van Rooij (following earlier accounts) imposes conditions like the following: rule out the case in which a counts as P in the context $\{a, b, c\}$ but a does not count as P in the context $\{a, b\}$. By imposing this condition and others like it, this choice rule generates a weak order among the P entities $\succsim$.[6] From this weak order, it is straightforward to derive a strict comparative relation $>$ (12a).

(12) a. Strict: $x > y$ iff $x \succsim y$ but not $y \succsim x$.

b. Indifference: $x \sim y$ iff $x \succsim y$ and $y \succsim x$.

However, if the choice rule R is interpreted as a complete and transitive weak order $\succsim$, then R will not model vagueness effects. This is because such a relation is transitive in its indifference relation $\sim$ (12b) and will validate the sorites paradox. For example, suppose $[\![\text{happy}]\!]^{\{x,y\}} = \{x, y\}$ and $[\![\text{happy}]\!]^{\{y,z\}} = \{y, z\}$. Then $x \sim y$, and $y \sim z$. By the transitivity of $\succsim$ (and hence $\sim$), $x \sim z$. This implies the conclusion

[6] In part for the reasons discussed below, van Rooij (2011) ultimately settles on a choice rule generating a semi-order.

of the sorites paradox. Since the sorites paradox is invalid, a choice rule generating a weak order fails to respect the semantics of vague predicates.

Thus, a semantic analysis of vague predicates seemingly requires a choice rule that gives rise to both a transitive strict relation and an intransitive indifference relation. The trouble arises with combinations of the relations, as in (13).

(13) $a \succ b$ and $b \sim c$

As Sen (1970) shows, if (13) implies that $a \succ c$, then the indifference relation is transitive—a result at odds with the sorites paradox. On the other hand, if (13) does not imply that $a \succ c$, then (13) implies that either $c \succ a$ or $a \sim c$. If the former, then the comparative relation is intransitive, which is contrary to most semantic analyses of comparatives. If the latter, then it is possible for two entities to differ with respect to property P but to remain unrelated by the comparative relation defined over P. For example, Sam and Sonia could both qualify as healthy and yet Sam could be neither healthier nor as healthy nor less healthy than Sonia. As discussed below, the latter option may be plausible for a class of adjectives called multidimensional adjectives.[7]

3 Multidimensional Adjectives as Collective Choice Functions

Whether Clarence is *healthy* intuitively depends on a number of factors simultaneously, like blood pressure, heart function, cholesterol levels, or more, depending on the context of use. Adjectives like *healthy* are associated with multiple dimensions (or criteria) of evaluation (Kamp 1975; Klein 1980; Sassoon 2013a, b). For example, *similar, identical, typical, normal, good, happy,* and *healthy* are all multidimensional adjectives.

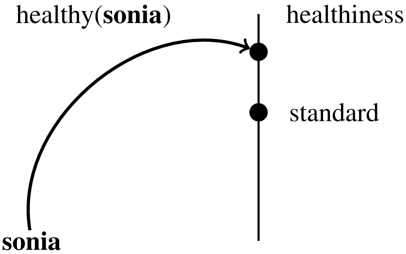
[7] It is also possible to start with a strict preference relation $\succ$ and to derive indifference and the relation $\succsim$ from there (1a).

(1) a. Indifference: $x \sim y$ iff $\neg x \succ y$ and $\neg y \succ x$
    b. Weak preference: $x \succsim y$ iff $\neg y \succ x$

Importantly, this rule gives rise to an intransitive indifference relation in some cases. In choosing between bundles of goods (like beer and wine), an individual may prefer 10 cans of beer and 2 bottles of wine to 9 cans of beer and 2 bottles of wine: $(10, 2) \succ (9, 2)$. However, such an individual may be indifferent as between these two options and 5 cans of beer and 3 bottles of wine: $(10, 2) \sim (5, 3)$ and $(9, 2) \sim (5, 3)$. Note, then, that both (9, 2) and (10, 2) are indifferent to (5, 3), but $(10, 2) \succ (9, 2)$. One of the latter two bundles is strictly preferred to the other, but both are indifferent to a third.

This entails the intransitivity of the indifference relation, and such intransitivity is plausible when the ranked entities have multiple attributes that figure into the ranking (like wine and beer). As discussed below, the choice rule adopted by theories of vagueness includes a transitive strict preference relation and an intransitive indifference relation along the lines of (1a). Such a choice rule suggests that the set of available options X is composed of bundles—that is, the choice rule is "multidimensional."

On standard semantic accounts, gradable adjectives like *healthy* are measure functions that map an entity to a position on a one-dimensional scale of healthiness, as in Fig. 14 (Kennedy 1999). These measure functions combine with a phonologically null morpheme to derive the denotation in (15).

(14) 

(15) $[\![\text{healthy}]\!] = \lambda x.\text{healthy}(x) \geq \mathbf{s}(\text{healthy})$

Paraphrased, x is healthy if x's health is greater than or equal to some "standard of comparison" relative to a set of contextually salient entities, called a "comparison class."[8] On this account, the degrees associated with gradable adjectives are formalized by a triple $\langle D, <, \delta\rangle$ including the set of degrees D, an ordering on this domain $<$, and a dimension δ that provides the property to be measured (for instance, cost in the case of *expensive*) (Kennedy and McNally 2005).

However, the property measured by an adjective like *healthy* appears to comprise many dimensions (Sassoon 2013a). For example, it is possible to specify a dimension overtly with prepositional phrases like *with respect to* or *in*, as in (16). Such specifications are not available for other adjectives like *long*. Despite having both temporal and spatial interpretations, a single utterance of *long* is not interpreted relative to both dimensions at the same time (16b).

(16) a. John is healthy with respect to blood pressure.
b. * The wedding is long {with respect to, in} temporal duration (but not with respect to space).

It is also possible to quantify over multiple dimensions, as in (17), and to target dimensions with *wh*-words (18).

(17) a. Elena is healthy in {every respect, some respects, most respects}.
b. * The table is long in {all, most, three, some} respects.

(18) In what respect is Elena healthy?

And it is possible to except dimensions from consideration, as in (19).

[8]The standard-setting function **s** "is a context-sensitive function that chooses a standard of comparison in such a way as to ensure that the objects that the positive form is true of 'stand out' in the context of utterance, relative to the kind of measurement that the adjective encodes" (Kennedy 2007, p. 17).

(19) a. Ruth is healthy except with respect to her cholesterol.
b. * The table is long except with respect to temporal duration.

These tests suggest that an adjective like *healthy* is sensitive to measures along multiple dimensions. A "dimension" may be an explicit and quantifiable semantic argument, as in *healthy in every respect*, but dimensions also appear in a number of constructions that lack a quantifiable dimension argument (*mathematically, Elena is clever*).

On this picture, it is impossible to fix a scale for multidimensional adjectives without reference to some procedure for integrating mutliple dimensions. As alluded to above, such a procedure is necessary to ensure that comparatives are transitive. Therefore, where Kennedy and McNally (2005) associate gradable adjectives with a set of degrees D, an ordering, and a dimension δ providing the property to be measured, I associate multidimensional adjectives with a set of degrees, an ordering, and a function that creates one scalar property from many, $\langle D, <, f(\delta_1, \ldots, \delta_n)\rangle$. The function f is thus a collective choice function.

Moreover, English grammar constrains how dimensional measures should affect the meaning of the multidimensional adjective. For instance, if every dimension involved in a use of *healthy* ranks Elena sufficiently highly, then a proper use of *healthy* should reflect this ranking. Sentences like (20) should therefore seem semantically odd.

(20) # Elena is healthy than in every respect, but she is not healthy.

The oddness of (20) reflects (one implication of) Arrow's principle of Unanimity: if every dimension ranks Elena as high or higher than other members of the comparison class, the aggregated ranking should also rank Elena as high or higher than those same members.

Similarly, it should be possible to determine the relative health of Nino and Ruth by considering only the relative rankings of Nino and Ruth along the relevant dimensions. A phenomenon known as "implicit comparison" may verify this effect (Kennedy 2011). Imagine a dialogue as in (21).

(21) Speaker A: Compared to Ruth, Nino is healthy.
Speaker B: But consider Clarence.
Speaker A: # Then, compared to Ruth, Nino is not healthy.

If Speaker A's conclusion in (21) is anomalous, then this is a linguistic analog of Independence of Irrelevant Alternatives.[9] In (21), Clarence is an irrelevant alternative.

We also have no reason to believe that some possible rankings of members of the comparison class should be excluded at the outset. I take this to support the initial plausibility of Unrestricted Domain.

An analog of Non-Dictatorship may be inferred from the oddness of (22).

[9] A reviewer points out that this version of Independence reflects a fixed ranking of alternatives, in contrast to the standard Arrowian context of many varying rankings of alternatives. Arrow's result does not turn on this distinction (Feldman and Serrano 2008), but it may raise questions about the strength of Unrestricted Domain in the semantics of gradable adjectives (Morreau 2015).

(22) (Scenario: Sonia has good blood pressure, but terrible cholesterol and heart rate.)
Sonia is healthy.

In (22), the speaker's conclusion is odd because one dimension does not usually dictate the meaning of a multidimensional adjective. If the dimension of blood pressure were "dictatorial" in the context represented by (22), the speaker's conclusion should strike us as natural. Similarly, the speaker's conclusion in (23) may be natural, despite the middling ranking of Sonia with respect to blood pressure.

(23) (Scenario: Sonia has middling blood pressure, but good cholesterol and heart rate)
Sonia is healthy.

Taken together, (22) and (23) suggest that one dimension does not usually control the interpretation of a multidimensional adjective in a "dictatorial" fashion.

If gradable adjectives are choice functions, then they are defined with respect to a choice rule.[10]

(24) $[\![\text{healthy}]\!]^c = \{x \mid \forall y \in c,\ xRy\}$

For multidimensional adjectives, this choice rule is a function of the choice rules used by the multidimensional adjective's component dimensions.

(25) $R_{\text{healthy}} = f(R_{\text{bp}}, R_{\text{cholesterol}}, R_{\text{heart rate}}, \ldots)$

Arrow (1950) and Chichilnisky (1982) tell us that, under certain assumptions, the mapping f may result in a choice rule that is intransitive and discontinuous. For example, it is easy to recreate Condorcet's paradox in the adjectival context, replacing voters with dimensions and candidates with members of the comparison class.

(26)
bp:	john	>	ruth	>	stephen
cholesterol:	stephen	>	john	>	ruth
heart rate:	ruth	>	stephen	>	john

Furthermore, multidimensional adjectives appear to follow analogues to the Arrowian assumptions Unrestricted domain, Unanimity, Independence of Irrelevant Alternatives, and Non-dictatorship/anonymity.[11] Therefore, choice rules associated

[10] This remains a first approximation, and in particular R cannot be a weak preference relation ($\succsim$), as van Rooij (2011) points out. R could be some rule like "indifferent with respect to entities above the contextual standard or ranked higher than entities below the contextual standard." Whatever the choice rule, however, if it adheres to the Arrowian assumptions (which themselves appear to be accurate constraints on the meaning of multidimensional adjectives), it remains subject to the limitations on collective choice described above.

[11] Arrow's approach depends on *ordinal* dimensional rankings. Ordinal scales are too informationally impoverished to allow numerical measurements like *two meters* or ratio measurements like *twice as*. Moltmann (2005) uses this as a basis for proposing an ordinal analysis for multidimensional adjectives, since adjectives like *clever* and *happy* are also incompatible with numerical and ratio modifiers (Sassoon 2013b).

with multidimensional adjectives like $R_{\text{healthy}} = f(R_{\text{bp}}, R_{\text{cholesterol}}, R_{\text{heart rate}}, \ldots)$ may be intransitive. And as Chichilnisky (1982) shows, if the choice rule f is unanimous and anonymous, it will be discontinuous.

4 Some Vagueness Effects May Follow from Limitations on Collective Choice

Arrow and Chichilnisky's results show that seemingly sensible constraints on f—like Unanimity, Unrestricted Domain, Independence of Irrelevant Alternatives, and Non-Dictatorship—may lead f to fail in particular ways. Either the constraints themselves will be violated or other desirable properties like transitivity will be violated. Vagueness effects like the sorites paradox and borderline cases flow from these failures.

First, the role of intransitivity in vagueness effects is well-explored by Cobreros et al. (2012), who claim that "the non-transitivity of the indifference relation is a central feature of all vague predicates" (p. 349). Cobreros et al. (2012) build a theory of vagueness upon such intransitivities, training their analysis on the second premise of the sorites paradox (27), which represents the idea that vague predicates are "tolerant."

(27) Premise 1: A person whose systolic blood pressure is 100 is healthy.
Premise 2: A person whose systolic blood pressure is one unit higher than a healthy person's is healthy.
Conclusion: A person whose systolic blood pressure is 180 is healthy.

More generally, a tolerant predicate P obeys (28) for "indistinguishable" entities x_i and x_{i+1}.

(28) $\forall x_i, x_{i+1} : (P(x_i) \wedge x_i \sim^P x_{i+1}) \rightarrow P(x_{i+1})$

For example, the relation "indifferent with respect to health" is reflexive and symmetric, but not transitive: Elena's health maybe observationally indistinguishable from Sonia's, and Sonia's may be observationally indistinguishable from Nino's, but Elena and Nino may have observationally distinct measures of healthiness.

Tolerance implies both the transitivity of the indifference relation and the idea that small changes along one dimension of a multidimensional predicate result in small changes in the denotation of the multidimensional predicate.[12] First, tolerance implies transitive indifference. For example, (28) implies that if $x_i \sim x_{i+1}$ and if $x_{i+1} \sim x_{i+2}$, then $x_i \sim x_{i+2}$ for all x_i. If *healthy* is tolerant—as the second premise of (27) appears to suggest—then the indifference relation associated with the choice rule R_{healthy} is transitive. However, if R_{healthy} is the product of a collective choice

[12]This follows assuming that P is non-trivial and monotonic with respect to the relevant strict ordering P-er.

function f (25), then limitations on collective choice entail that f might not result in a transitive collective choice. The indifference relation $\sim_{\text{healthy}}$ may therefore be intransitive. Since tolerance implies the transitivity of the indifference relation, and since collective choice shows that this relation might not be transitive, tolerance is false.

Indeed, it is possible to trace this intransitivity to a particular Arrowian axiom. Independence of Irrelevant Alternatives restricts the collective choice function f in ways that give rise to intransitivity. As Saari (2008, p. 44) notes, if the aggregation function f obeys Independence, f may be rewritten as (29), where each component of f is determined by f's ranking of a pair of entities.[13]

(29) $f = (f_{\{\text{Elena, Sonia}\}}, f_{\{\text{Sonia, Nino}\}}, f_{\{\text{Nino, Elena}\}})$

It is easy to see that such a rule allows for intransitive outcomes: f in (29) can rank Elena indifferent to Sonia, Sonia indifferent to Nino, and Nino over Elena, without contradiction.

Fara's (2002) Similarity Constraint embodies the principle of Independence. In order to derive vagueness effects, Fara (2002, p. 59) relies on the Similarity Constraint in (30).

(30) **Similarity Constraint**. Whatever standard is in use for a vague expression, anything that is saliently similar, in the relevant respect, to something that meets the standard itself meets the standards; anything saliently similar to something that fails to meet the standard itself fails to meet the standard.

The way the Similarity Constraint works in practice, as Fara explains with respect to the vague predicate *tall*, is the following: "for any particular x and y that differ in height by just 1 mm, the very act of our evaluation raises the similarity of the pair to salience, which has the effect of rendering *true* the very instance we are considering." The Similarity Constraint requires "salient" comparisons, which turns out to mean pairwise comparisons. Fara rules out the universally quantified tolerance principle (28) because each comparison cannot be simultaneously salient to the speaker: "there are too many pairs for us to actively entertain each similarity" (p. 69).

Second, tolerance implies continuity.[14] If the tolerance principle in (28) is true, then small changes along one dimension of health, like blood pressure, result in

[13] Saari (2008) explains: "[Independence] requires that when determining the {Sibelius, Beethoven} societal ranking, the decision rule *cannot* use any information about how the voters rank other pairs: As such, this requirement means that *the rule cannot use any information about whether the voter have, or do not have, transitive preferences.*"

[14] The definition of continuity is dependent on the particular topology imposed on the choice space (that is, what constitutes the "open" sets).

(1) **Continuous map**: Let X and Y be topological spaces. A map $f : X \rightarrow Y$ is called continuous if the inverse image of open sets is always open. Jänich and Levy (1995)

"[A]ny reasonable topology, certainly any topology which has been used on preference spaces" reduces to a Euclidean choice space (Heal 1997, p. 3) in which continuity may be defined as follows:

small changes in the value of *healthy*. In particular, where x_i and x_{i+1} differ only by one unit of systolic blood pressure, either both are healthy ($[\![\text{healthy}(x_i)]\!] = 1 = [\![\text{healthy}(x_{i+1})]\!]$) or both are not ($[\![\text{healthy}(x_i)]\!] = 0 = [\![\text{healthy}(x_{i+1})]\!]$). However, R_{healthy} is the product of a collective choice function f (25), and limitations on collective choice entail that f might not result in a continuous collective choice. Since tolerance implies a continuous relationship between dimensional measures and the meaning of a multidimensional predicate, and since collective choice shows that this relation might not be continuous, tolerance is false.

Importantly, Smith (2008) notes that continuity is too strong a constraint for modeling tolerance. Continuity overgenerates the number of vague predicates in part because the definition of continuity is so flexible.[15] Topologically speaking, every discrete domain could support a continuous function, but this elides important distinctions between vague predicates like *healthy* and non-vague predicates like *has 100 hairs or fewer.* Thus, Smith adopts a "local" notion of continuity that is equivalent to the intuitive description of continuity I have been using: a function is locally continuous if a *small* change in input produces at most a *small* change in the value of the function.

Weber and Colyvan (2010) echo this tension between a "local" and "global" interpretation of tolerance in putting forth a topological version of the sorites paradox. According to Weber and Colyvan (2010), "[a] predicate is vague iff its characteristic function is locally constant but not globally constant" (p. 318), where local constancy is understood in a way similar to "local" continuity.[16]

Both Smith (2008) and Weber and Colyvan (2010) identify the seeming incompatibility between a vague predicate and the space in which that vague predicate lives. Further, both identify this incompatibility in terms of a distinction between the "local" and "global" properties of vague predicates. Collective choice theory provides one explanation of this distinction: incremental movements along one dimension of a vague predicate (the local level, like blood pressure in the case of *healthy*) do not

---

(2) **Continuity**: for any map $f : \mathbb{R}^m \to \mathbb{R}^k$, the map f is continuous at $p \in \mathbb{R}^m$ if for each $\epsilon > 0$, there exists a $\delta > 0$ such that $|f(q) - f(p)| < \epsilon$ as soon as $|q - p| < \delta$.

[15] In order to define continuity, we need to impose a topology on the set constituting the domain of the predicate P. For example, if a subset S of the domain satisfies the following property, S is a basis element of the relevant topology.

(1) $(x \leq_z y \wedge y \in S \wedge z \in S) \to x \in S$

That is, if x is at least as close to z as y is, and y and z are both in S, then x is in S. (This is Smith three-place similarity relationship, but he makes a convincing argument that a two-place similarity relationship works similarly.) Then $\forall x, y (x \leq_y x)$, which means that every singleton set is a basis element of the topology, and, consequently, every function between sets of entities is continuous. Such a topology fails to discriminate between linguistically vague predicates and predicates like *has 100 hairs or fewer.* The latter is not vague. Yet if we impose the topology generated by (1) on the domains of all predicates, all predicates will turn out to be vague: an unwelcome result.

[16] As the authors write, "[s]ubstituting 'continuous' for 'constant'…would make no great difference," (Footnote 17) acknowledging Smith objections to global continuity.

necessarily map into incremental movements along the semantic scale of the vague predicate itself (the global level, the scale of healthiness in the case of *healthy*).

Saari (2008, p. 13) too, understands pathologies of collective choice in a multidimensional choice space as a result of the tension between local and global properties. Collective choice functions lose information when mapping multiple "local" dimensional rankings (like the ranking of entities according to their blood pressure) onto a "global" aggregated ranking (like the scale of healthiness). This lost information derives from the effect of Independence—if the aggregation function f of a multidimensional adjective obeys Independence, then the aggregation function is equivalent to one that considers only pairs of entities—and also, Saari argues, from the effect of continuity. Continuity is itself a local concept that "ignores" global structure, and "this feature of ignoring the global structure is the same local versus global problem that arises with" Independence (p. 57).

This "local" versus "global" dichotomy also pervades other accounts of vagueness phenomena. In Cobreros et al. (2012) many-valued approach, for example, there are three notions of truth: classical truth, tolerant truth, and strict truth. Tolerant truth is a weaker notion of truth than classical truth, and strict truth is a stronger notion. The "local" inferences—the inferences that make up the second premise of the sorites argument—are tolerantly true. The idea is that you can only string so many tolerantly true inferences together (at the same time) before you end up with a false conclusion.[17] For example, Cobreros et al. (2012, p. 376) validate each individual step in the sorites premise (31a), but not the universally-quantified version (31b).[18]

(31) a. Premise 1: $P(a)$
Premise 2: $a \sim b$
Conclusion: $P(b)$

b. Premise 1: $P(a)$
Premise 2: $\forall i \in [i, n-1]\; a_i \sim a_{i+1}$
Conclusion: $P(n)$

Similarly, contextualists like Fara (2002) affirm each instance of the sorites premise (*if* n *is* P, *then* $n+1$ *is* P) without affirming the universally-quantified version ($\forall n$ *if* n *is* P, *then* $n+1$ *is* P). As Smith (2008, p. 113) describes contextualist accounts, "the basic idea behind contextualism is that vagueness is a diachronic phenomenon, which only emerges when we consider the semantic state of a language

[17]Other many-valued logics share this feature. For example, infinitely-valued logics often use a value along the interval [0, 1] as the "degree of truth" of a certain proposition (Zadeh 1975; Hyde 2008). Infinitely-valued logics are often defined in such a way that an argument is only valid if it preserves the maximum degree of truth in the argument's premises. The sorites argument fails to be valid on this account because each step in the second premise, while "almost true" itself, reduces the degree of truth of the conclusion bit by bit. In this way, the first premise is perfectly true, each instantiation of the tolerance principle is *almost* perfectly true, and the conclusion is perfectly false. Since the sorites argument fails to preserve the maximum degree of truth, it is invalid.

[18]The key to the validity of (31a) and the simultaneous invalidity of (31b) is an intransitive notion of logical consequence.

over time (or more generally, over multiple instances of interpretation).” Contextualist accounts place particular emphasis on the act of interpretation, arguing that the act itself changes the semantic “facts on the ground.” For example, Fara (2002, p. 59) explains that “for any particular x and y that differ in height by just 1 mm, the very act of our evaluation raises the similarity of the pair to salience, which has the effect of rendering *true* the very instance we are considering.” Tellingly, however, if we consider each instance of the sorites premise simultaneously, the sorites paradox resurfaces. Fara’s response is that such simultaneous consideration is impossible. Each comparison cannot be simultaneously *salient* to the speaker: “there are too many pairs for us to actively entertain each similarity” (p. 69).

The social choice approach identifies the “local” versus “global” dichotomy as the paradox of choice in multidimensional space. Tolerance suggests that comparisons of the type that form the second premise of the sorites paradox take place along a scale that looks well-behaved (that is, is transitive and continuous). However, because the scale is actually is multidimensional, it is not well-behaved. As Saari (2008, p. 57) explains, “if we lived on a huge circle and used only local information, we would view the circle as being, essentially, a straight line. It is not a line; it is a circle.” Arrow and Chichilnisky’s results explain why failures of transitivity and continuity might arise–and hence why vagueness effects might arise–at least where multidimensional adjectives are understood as collective choice functions subject to certain assumptions.

However, it does not follow that every collective choice invokes intransitivity or discontinuity. In order to fully honor Lewis’s claim that vagueness is semantic indecision, I need to demonstrate that the domain of dimensional rankings is dense enough and diverse enough to generate social choice paradoxes (Gaertner 2001). While there is a strong combinatoric argument for this (Saari 2008), I leave this justification for future work. Nor is what I have said here straightforwardly applicable to unidimensional adjectives like *tall*.

Social choice theory makes clear the sense in which vagueness effects may result from paradoxes of decision making. On this view, the sorites paradox and Condorcet’s paradox derive from the same source: it is impossible to make a collective decision over a multidimensional choice space under reasonable constraints. So we waffle; we flip-flop; we vacillate. In other words, vagueness is semantic indecision.

References

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *The Journal of Political Economy*, *58*, 328–346. https://doi.org/10.1086/256963.

Baryshnikov, Y. M. (1997). Topological and discrete social choice: in a search of a theory. In G. Heal (Ed.), *Topological social choice* (pp. 53–63). Berlin: Springer. https://doi.org/10.1007/978-3-642-60891-9-3.

Chichilnisky, G. (1982). Social aggregation rules and continuity. *The Quarterly Journal of Economics*, *97*(2), 337–352. https://doi.org/10.2307/1880762.

Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2012). Tolerant, classical, strict. *Journal of Philosophical Logic*, *41*(2), 347–385. https://doi.org/10.1007/s10992-010-9165-z.

Fara, D. G. (2002). Shifting sands: an interest- relative theory of vagueness. *Philosophical Topics*, *28*(1), 45–82. https://doi.org/10.5840/philtopics20002816. originally published under the name Delia Graff.

Feldman, A,, & Serrano, R. (2008). Arrow's impossibility theorem: Two simple single-profile versions. Brown University Department of Economics Working Paper No 2006–11

Gaertner, W. (2001). *Domain conditions in social choice theory*. Cambridge: Cambridge University Press.

Gaertner, W. (2009). *A primer in social choice theory*. Oxford: Oxford University Press.

Grinsell, T. W. (2014). Unlikely imperfectives. In T. Snider, S. D'Antonio, & M. Weigand (Eds.), *Semantics and linguistic theory* (Vol. 24, pp. 395–413). https://doi.org/10.3765/salt.v24i0.2431.

Heal, G. (1997). Social choice and resource allocation: A topological perspective. In G. Heal (Ed.), *Topological social choice* (pp. 1–14). Dordrecht: Springer. https://doi.org/10.1007/978-3-642-60891-9_1.

Hyde, D. (2008). *Vagueness, logic and ontology*. London: Routeledge.

Jänich, K., & Levy, S. (1995). *Topology*., Undergraduate texts in mathematics New York: Springer.

Kamp, H. (1975). Two theories about adjectives. In E. Keenan (Ed.), *Formal semantics of natural language* (pp. 123–155). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511897696.011.

Kamp, H. (1981). A theory of truth and semantic representation. In P. Portner & B. H. Partee (Eds.), *Formal semantics: the essential readings* (pp. 189–222). Oxford: Blackwell. https://doi.org/10.1002/9780470758335.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*(1), 1–45. https://doi.org/10.1007/s10988-006-9008-0.

Kennedy, C. (2011). Vagueness and comparison. In P. Égré & N. W. Klinedinst (Eds.), *Vagueness and language use* (pp. 73–97). London: Palgrave Macmillan.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. In *Language*, *181*(2), 345–381. https://doi.org/10.1353/lan.2005.0071.

Kennedy, C. (1999). *Projecting the adjective: the syntax and semantics of gradability and comparison*. New York: Routledge.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*(1), 1–45. https://doi.org/10.1007/BF00351812.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). *Microeconomic theory* (Vol. 1). Oxford: Oxford University Press.

Moltmann, F. (2005). Comparatives without degrees: A new approach. In *Proceedings of the fifteenth amsterdam colloquium* (pp. 155–160).

Morreau, M. (2015). Theory choice and social choice: Kuhn vindicated. *Mind*, *124*(493), 239–262. https://doi.org/10.1093/mind/fzu176.

Saari, D. G. (1997). Informational geometry of social choice. In G. Heal (Ed.), *Topological social choice* (pp. 65–86). Berlin: Springer. https://doi.org/10.1007/978-3-642-60891-9_4.

Saari, D. G. (2008). *Disposing dictators, demystifying voting paradoxes*. Cambridge: Cambridge University Press.

Sassoon, G. W. (2013a). A typology of multidimensional adjectives. *Journal of Semantics*, *30*(3), 335–380. https://doi.org/10.1093/jos/ffs012.

Sassoon, G. W. (2013b). *Vagueness, gradability and typicality: The interpretation of adjectives and nouns*. Leiden: Brill. https://doi.org/10.1163/9789004248588.

Sen, A. K. (1970). *Collective choice and social welfare*. Netherlands: North-Holland.

Smith, N. J. (2008). *Vagueness and degrees of truth*. Oxford: Oxford University Press.

van Rooij, R. (2011). Vagueness and linguistics. In G. Ronzitti (Ed.), *Vagueness: A guide* (pp. 123–170). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-0375-9.

Weber, Z., & Colyvan, M. (2010). A topological sorites. *The Journal of Philosophy*, *107*(6), 311–325. https://doi.org/10.5840/jphil2010107624.
Williamson, T. (1994). *Vagueness*. London: Routledge.
Zadeh, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthese*, *30*(3–4), 407–428. https://doi.org/10.1007/BF00485052.

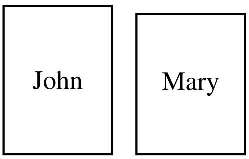
Grounding a Pragmatic Theory of Vagueness on Experimental Data: Semi-orders and Weber's Law

Robert van Rooij and Arnold Kochari

Abstract One of the traditional pragmatic approaches to vagueness suggests that there needs to be a significant gap between individuals or objects that can be described using a vague adjective like *tall* and those that cannot. In contrast, intuitively, an explicit comparative like *taller* does not require fulfillment of the gap requirement. Our starting point for this paper is the consideration that people cannot make precise measures under time pressure and their ability to discriminate approximate heights (or other values) obeys Weber's law. We formulate and experimentally test three hypotheses relating to the difference between positive and comparative forms of the vague adjectives, gap requirement, and Weber's law. In two experiments, participants judged appropriateness of usage of positive and comparative forms of vague adjectives in a sentence-picture verification task. Consequently, we review formal analysis of vagueness using weak orders and semi-orders and suggest adjustments based on the experimental results and properties of Weber's law.

1 Introduction

Indirect versus direct comparisons. Consider the following figure, depicting the heights of John and Mary.



Thanks to the Dutch NWO gravitation project *Language in Interaction*, the ESSENCE Marie Curie Initial Training Network, and the project 'Logicas no-transitivas. Una nueva aproximacion a las paradojas' for financially supporting this project. The authors also thank Nicole Standen-Mills for professional proof-reading.

R. van Rooij (✉) · A. Kochari
Institute for Logic, Language and Computation, Amsterdam, The Netherlands
e-mail: r.a.m.vanrooij@uva.nl

R. Dietz (ed.), *Vagueness and Rationality in Language Use and Cognition*,
Language, Cognition, and Mind 5, https://doi.org/10.1007/978-3-030-15931-3_9

According to Kennedy's (2011) intuition, this picture allows us to say (1a). At the same time, his intuition is that (1b) is false.

(1) a. John is taller than Mary.
b. Compared to Mary, John is tall, but compared to John, Mary is not tall.

Kennedy proposes that whereas for the truth of the *explicit* comparative (1a) any (directly or indirectly observable) bit of John's height that exceeds Mary's suffices, there should be a *significant gap* between the heights of John and Mary for the *implicit* comparative (1b) to be true. According to Kennedy (2011), such a significant gap is also required to make the positive sentence (2) true with respect to an implicit comparison class:

(2) John is tall.

As it turns out, Kennedy's intuition concerning (2) fits well with a pragmatic approach towards clustering proposed by various philosophers to (dis)solve the Sorites paradox.

The pragmatic gap requirement We can divide people into those that are tall and those that are not. How do we do so? What is the criterion? According to the most obvious solution, everybody is tall whose height is above a certain cutoff-point, where every bit of height suffices. But that gives rise to several questions: (i) How do we determine the cutoff-point? Is it the mean height of the relevant persons, the median height, or something else?[1] (ii) How does this proposal account for the intuition that some people are borderline cases of tall people? (iii) Doesn't this presuppose that we need to be able to measure people's height before we can classify people as tall, and isn't that problematic? Suppose we have settled question (i). How, then to account for borderline cases? A natural proposal would have it that those individuals are borderline tall for which we cannot determine whether they are above or below the cutoff-point. This could be because our measurement is in many cases imprecise. But once we allow for this, we immediately get into the well-known *Sorites problem*. In a Sorites problem, we have a long series of people ordered in terms of their height, where we assume that the difference in height between two subsequent persons is always very small. Now, if you decide that the first individual presented to you, the tallest, is tall, it seems only reasonable to judge the second individual to be tall as well, since there is only a miniscule difference in heights. But, then, by the same token, the third person must be tall as well, and so on indefinitely. In particular, this makes the last person also tall, which is a counterintuitive conclusion given that it is in contradiction with our intuition that this last, and shortest individual, is short, and thus not tall.

According to two pragmatic approaches towards vagueness, the problem of borderline cases is more of a theoretical than a practical problem. In practice—or so

[1] For experimental research on this, see Schmidt et al. (2008) and Solt and Gotzner (2012). Solt and Gotzner (2012), for instance, show that classification is based on more than just ranking.

these pragmatic approaches claim—we are hardly ever confronted with the problem of borderline cases, because we have found pragmatic strategies to avoid them.

According to one such a pragmatic proposal, as suggested by Fara (2000) and foreshadowed by Tafel and Wilkens (1963), our clustering of classes into tall and non-tall objects can make use of any salient feature that distinguishes one group from the other, as long as this clustering is roughly consistent with a grouping based on height. One such salient feature in terms of which one group is distinguished from another might be the *spatial arrangement* of objects.

To examine this proposal, we are testing in an ongoing series of experiments whether a specific arrangement of objects matters for the applicability of a vague adjective as a description of one of these objects. If it is indeed the case that any salient feature can be used for clustering for the purpose of vague adjective applicability, then spatial arrangement could be one good candidate for such a feature. It would predict that when the same object visually appears to be grouped with bigger objects, it should be classified as *big*, and when it visually appears to be grouped with smaller objects, it should be classified as *non-big*. In our experiments, we presented participants with 6 (or 11 in one of the experiments) objects standing next to each other. These objects were equally spaced from each other in terms of the relevant property (specifically, we used size, length, height, and width) and the target object was in the middle of this distribution. We created different conditions by manipulating the order in which the objects were arranged. The participants' task was to decide whether the target object can be described as *tall/big/long/high*. In several experiments that we have conducted so far, the order did not matter for applicability of the vague adjective. Our intermediate results presented here informally thus suggest that there is some kind of criterion of applicability of vague adjective which can not be overridden/substituted by the visual grouping of the objects. Although this somewhat opportunistic pragmatic hypothesis may seem plausible, so far we have failed to find experimental evidence in support of this proposal.

A second pragmatic approach to clustering has explicitly been proposed as a solution to the Sorites paradox (Gaifman 2010; Pagin 2010, 2011; Rayo 2010; van Rooij 2011b, arguably all based on Wittgenstein 1953). According to this hypothesis, we can appropriately classify some member of a group as being tall only if the clustering of the group based solely on height does not give rise to borderline cases. This is the case precisely if there is a significant gap in height between the shortest person classified as tall, on the one hand, and the tallest person of the other group, on the other. According to this *gap hypothesis*, x cannot be classified as tall if there is another person y not classified as tall, but still similar in height, $x \sim_T y$. This solution seems natural: we make a useful division of the set of all relevant objects into those that are tall and those that are not, and this division is easy if there exists a significant gap. This approach to clustering solves—or perhaps dissolves[2]—the Sorites paradox, because the tolerance principle, according to which if one person is tall, and another has a very similar height, this second person is tall as well, i.e., $\forall x, y((Tx \wedge x \sim_T y) \rightarrow Ty)$, will not give rise to inconsistency.

[2] According to this approach, the Sorites paradox dissolves, because it does not even arise.

Analogue magnitude representation and Weber's law In the discussion so far, it seems that we have implicitly assumed that we need to be able to *measure* people's height in order to classify people as tall, or taller than another. But there is serious doubt that these and other classifications really require an ability to measure persons, objects, and events, in terms of meters, liters, hours and other units of precise measurements with which we are so familiar, making use of a mathematically defined system of numbers. Psychological evidence (starting from Moyer 1973) suggests that people represent these magnitudes in a more automatic, but imprecise (i.e., approximate), analogue way, and that we use them in this way, for instance, under time-pressure and other cognitive loaded circumstances.[3] These so-called Analogue Magnitude Representations (AMRs) are taken to be primitive because in contrast to the standard way we think about measures, they are already present in infants, as well as in some other animals, such as primates, dogs, and birds.

One important feature of AMRs is that they obey *Weber's law*, which holds that the ability to discriminate two magnitudes is determined by their ratio, i.e., that this ability grows with intensity (Weber 1834).[4] According to this law, it is more difficult to discriminate two stimuli of high intensity than it is to discriminate two stimuli of low intensity, even though the absolute difference between the intensity of two high stimuli is exactly the same. As suggested above, AMRs exist for instance for numerical magnitudes, via our so-called 'approximate number system' (Dehaene 2011). And indeed, experimental data (e.g. Dehaene 2011) confirms that under natural circumstances the approximate discrimination between the cardinality of two sets can only be done when they differ by a certain percentage. A natural hypothesis is that we use AMRs for all types of magnitudes, also including, for instance, length/height. But this also suggests that when we compare individuals in terms of their approximate height, our ability to distinguish these heights, at least when we can not measure them precisely, will obey Weber's law. Applying what we know about AMRs and the principles of Weber's law in terms of perception to vague adjectives, we would expect that there too, the higher the relative difference between objects, the more applicable a vague predicate would be as a description.[5] This is in accordance with the *gap hypothesis* and can be seen as its extension.

Three hypotheses The above discussion of the difference between explicit versus implicit comparatives, of the requirement of a gap and the magnitude thereof for the use of positive sentences, based on the idea that we represent magnitudes in

[3]The different circumstances in which the different representations are used brings to mind the so-called Dual Process Theories of reasoning (cf. Evans 2010; Kahneman 2011), according to which there that are two distinct processing modes available for many cognitive tasks, one being automatic and fast, while the other is more controlled and slower. We will not dwell on this more general hypothesis in this paper, however.

[4]Strictly speaking, Weber's law is about relative *changes* in stimulus magnitude, rather than relative *differences* between two stimuli. Thus Weber's law on its own only says that the perceived change in a magnitude is proportional to relative changes in that magnitude. In this paper we assume that the likelihood of detecting a change in magnitude is proportional to the amount in perceived change, so that Weber's law as interpreted strictly comes down to our interpretation.

[5]See Égré (2017) for similar reasoning.

an analogue way, gives rise to the following three hypotheses for one-dimensional adjectives:

1. *Comparative hypothesis*: For an explicit comparative of the form ‘John is taller than Mary’ to be true, there does not have to be a significant gap in height between John and Mary, so any difference would suffice.
2. *Gap hypothesis*: The truthful use of a sentence with a positive adjective like ‘John is tall’ requires there to be a *significant gap* in height between John and other people within the set of relevant persons.
3. *Weber’s law hypothesis*: Acceptability of a sentence with a positive adjective like ‘John is tall’ will increase with increasing relative difference between the heights of John and other people within the set of relevant persons.

In the following section we will discuss to what extent these hypotheses are confirmed by experimental evidence. After that, we will look into the formal analysis of vagueness in terms of weak orders and semi-orders and address challenges to it based on the experimental evidence and AMRs.

2 Experimental Evidence

In this section, we present the results of two experiments on how vague adjectives are used by people when they compare a number of objects. In order to test the hypotheses presented above, we collected data in a sentence-picture verification task. In this task, participants read sentences describing an object using a vague adjective. They then saw a picture depicting this object and several others. The task was to decide whether the sentence was suitable as a valid description of the target object. We manipulated the relative differences in size between the target and other objects within the same image and expected that this would affect the acceptance rates.

Each presented picture contained 4 objects. The target object was of a different color than the other 3 comparison objects. This color was used to refer to this object in the sentences. The target object was always the biggest/longest/tallest/highest one in the picture. The comparison objects were smaller than the target object. They were all approximately of the same size, with a small, but noticeable difference. We manipulated the size difference between the target object and the biggest of the comparison objects. The ratio between the target and the biggest comparison object was one of the following: 12:12 (same size), 11:12, 9:12 or 7:12.

Participants had to read and make judgments about the same pictures in two conditions: once for a sentence describing the target using a positive form of a vague adjective and once for a similar sentence describing the target using a comparative form of this adjective. By comparing acceptance rates for the comparative and positive forms, we were able to see if people have different criteria for them. In sum, we had the following design: adjective form (positive versus comparative) X difference ratio (12:12, 11:12, 9:12 and 7:12).

We expected that sentences in combination with ratio 12:12 would never be accepted since it should be a factually false statement for positive and comparative forms. Based on our *comparative hypothesis*, we expected the sentence with a comparative form to be accepted whenever participants saw the difference at all, i.e. in all ratios except for 12:12. Based on our *gap hypothesis* and *Weber's law hypothesis*, we expected the acceptance rates of sentences with positive forms of adjectives to be different for different ratios since a small gap may not be sufficient. Thus, the positive form of a vague adjective should be more acceptable when the difference ratio is 7:12 or 9:12 than when it is 11:12. Ideally, we should observe the following pattern of acceptance rates: 11:12< 9:12< 7:12.

Besides the acceptance rates, we also collected reaction times, i.e. the amount of time participants took to make a decision. Reaction times (RTs) are conventionally considered to reflect the difficulty/complexity of the decision being made. Consequently, if there are more processes taking place or there is more hesitation in a trial, the reaction time for this trial should be longer. Thus, based on our hypotheses, for positive-form judgment, in case of lower ratios the decision should take longer, because a positive vague adjective is less applicable there. The reaction times should thus form the following pattern 11:12>9:12>7:12. In addition, based on our *comparative hypothesis* we expected that there would be a difference in reaction times of acceptance between comparative and positive adjectives. Positive adjectives should take longer because participants need not only to visually register the difference, but also make sure that whatever criterion they use for applicability of positive adjectives is met.

From our pilot experiments we noticed that intermixing sentences with positive and comparative forms of the adjectives can confuse participants, so we decided to first present all the sentences with positive adjectives (where participants can stick to their idea of what *tall, high,* etc., means) and then present all sentences with comparative adjectives. During the first half of the experiment, the participants did not know that they would be judging comparative forms in the second half, and they were never told that they would see the same pictures twice.

2.1 Experiment 1

Materials We used 4 different adjectives and there were 12 different objects in total. The following adjective and object pairs were used: *big*: box, apple, square, circle; *long*: pencil, line, drinking straw; *tall*: bar, can, building, street lamp; *high*: stool. The colors were chosen arbitrarily, though in such a way that they would not be confused with each other based on the subjective judgment of the experimenter.

The sentences with positive forms of adjectives were always constructed in the following way: 'The [target object color] [target object] is [adjective in positive form].', e.g. 'The grey box is big.' Sentences with the comparative forms were constructed in the following way: 'The [target object color] [target object] is [adjective in comparative form] than the [other color] ones', e.g. 'The grey box is bigger than

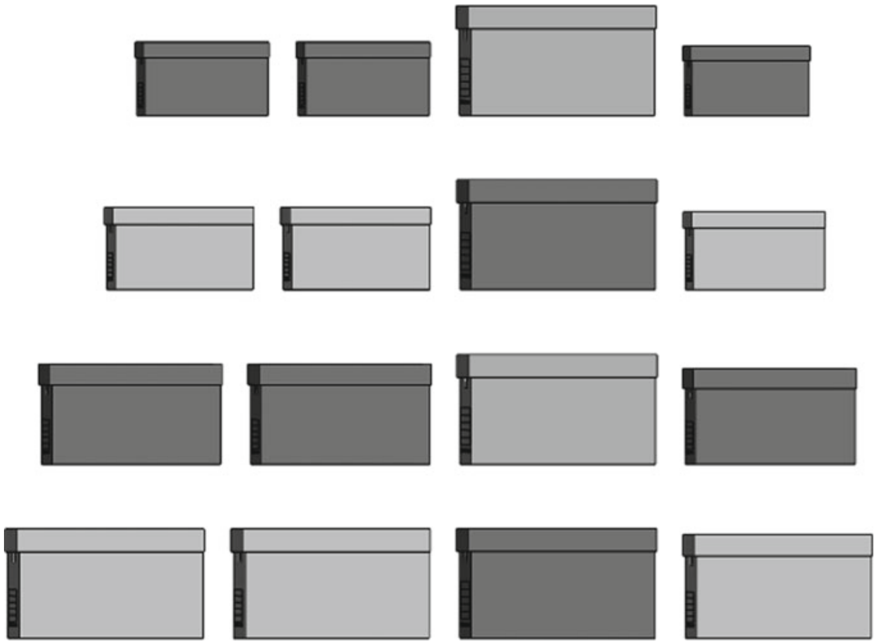


Fig. 1 Examples of stimuli used in Experiment 1. From top to bottom, ratios: 7:12, 9:12, 11:12, 12:12. Sentences: 'The grey/orange box is big' or 'The grey/orange box is bigger than the orange/grey ones'. Orange color here printed as dark grey

the orange ones.' The pictures used with these sentences are given as examples in Fig. 1.

Each of the twelve objects was presented in four ratios and in two adjective forms, so we had 12*2*4 = 96 experimental trials in total. In addition to the experimental trials, we added 8 filler trials, each of which was also presented once in the positive and once in the comparative form, so there were 16 additional trials. These filler trials used different objects, but were constructed in such a way that they looked exactly like the experimental ones except that the answer was always 'no', because the target object was one of the smaller ones and thus the statements were factually wrong. The fillers were intermixed with the other trials and should not have been noticed as different.

A detailed description of the experimental procedure is provided in Appendix A.

Participants Participants were recruited via Prolific.ac, an online platform aimed at connecting researchers and participants willing to fill in surveys and questionnaires in exchange for compensation for their time. We recruited native American English speakers. Twenty-four participants completed the task in order and were included in the analyses reported below. The detailed eligibility and exclusion criteria are provided in Appendix A.

Table 1 Mean acceptance rates per ratio and sentence type, Experiment 1

Ratio	Sentence type	Total N trials	% of YES responses
7:12	Positive	285	99
	Comparative	285	100
9:12	Positive	287	99
	Comparative	286	100
11:12	Positive	283	97
	Comparative	287	96
12:12	Positive	272	37
	Comparative	282	0
Fillers	Positive	190	1
	Comparative	191	1

Results Each of the 24 participants completed 112 trials, for a total of 2688 trials. We excluded all trials to which participants did not give a response (missing responses, N = 39) and all trials with RTs lower than 250 ms (N = 1) since it would be impossible to make a decision and to send a signal to execute the motor command (press the button) in such a short time. So in total we excluded 40 trials, or 1.5% of the total number of trials.

The acceptance rates for each ratio per sentence type are presented in Table 1 below. There seems to be no difference between the sentences with positive and comparative adjectives except for the ratio 12:12. In this latter ratio the target object was not the biggest anymore and, while 37% of participants still agreed that it could be referred to as *big/tall* etc, no participants agreed with the false statement that it was *bigger/taller* than other objects.

Besides the acceptance rates, we also expected to see differences in reaction times between sentences with different forms of vague adjectives. The reaction times per ratio and sentence type are shown in Fig. 2. In this and further reaction time analyses, we only look at reaction times to 'yes' responses, because 'no' responses are the result of a different decision and are not conventionally analyzed together with 'yes' responses. We can see that there is no difference between positive and comparative sentences except for the ratio 11:12, for which comparative sentence verification took slightly longer.

Discussion In this experiment, there was no difference between how participants judged the applicability of positive and comparative forms of vague adjectives for any of the ratios in which we were interested. For any ratio except 12:12, the participants agreed that both the positive form and the comparative form were applicable. There was no difference between individual participants and adjectives either. This was an unexpected result for us.

One case in which we would not see any difference between adjective forms would be if participants took the positive forms *tall, long, big, high* to be acceptable when the target object was simply the biggest/longest etc. If they interpreted a sentence like

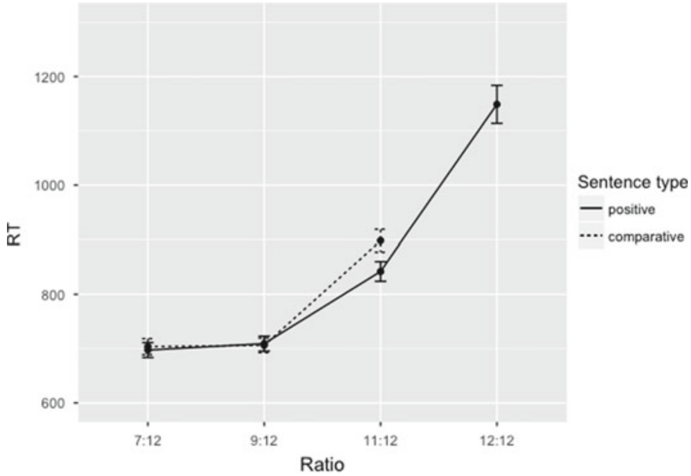


Fig. 2 Mean RTs per ratio and sentence type for 'yes' responses, Experiment 1. The value for the comparative sentences in ratio 12:12 is not shown since there was only one data-point. The error bars represent the standard error value

'The grey box is big' as meaning that the grey box is the biggest one, then indeed, we would expect a 'yes' response in all trials. In order to see if this would be the case, we conducted another experiment where everything was kept the same except that the target object was no longer the biggest/tallest/highest/longest, so the superlative interpretation would no longer be possible.

2.2 Experiment 2

In this experiment, we tested the same hypotheses as in Experiment 1, using the same set-up, but we added one more object to each of the pictures. If the positive forms were indeed interpreted as referring to the biggest, etc., object, then by making the target object not the biggest in the group of objects to be judged, we expected to obtain different results, potentially agreeing with our hypothesis.

Materials We added one more object to each picture presented in this experiment. This new object was bigger than the target object, but of the same color as the comparison objects. Specifically, it was always 110% of the size of the target object. Examples of altered stimuli can be seen in Fig. 3.

In addition, we also had to change the formulation of the sentences in the second block where participants saw comparative forms of the adjectives The new formulation was 'The [target object color] [target object] is [adjective in comparative form] than the [the other color] ones on its two sides', e.g. 'The grey box is bigger than the orange ones on its two sides.' The formulation had to be such in order for the

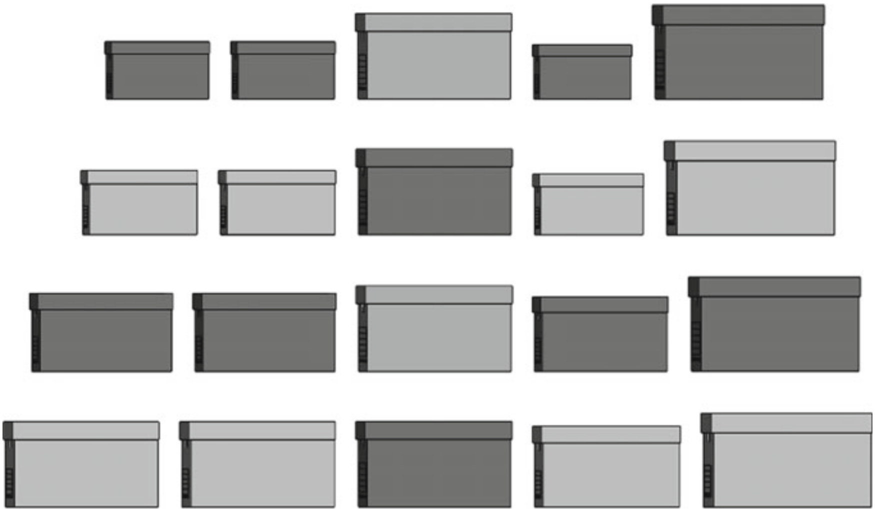


Fig. 3 Examples of stimuli used in Experiment 2. From top to bottom, ratios: 7:12, 9:12, 11:12, 12:12. Sentences: 'The grey/orange box is big' or 'The grey/orange box is bigger than the orange/grey ones on its two sides'. Orange color here printed as dark grey

sentences to still test whether the participants can visually register the difference. In order for this sentence to be true, the target object was never standing next to the newly added biggest object.

Participants The same eligibility requirements as in Experiment 1 were applied to participants in this experiment. Thirty-one participant completed the task in order and was included in the analyses. See Appendix A for details.

Results Each of the 31 participants included in the analysis did 112 trials, for a total of 3472 trials. We excluded all trials to which participants did not give a response (N = 58) and trials with RTs lower than 250 ms (N = 2). In total we excluded 60 trials or 1.73% of the total number of trials.

Data analyses reported here were conducted in the R environment (R Development Core Team 2016). In order to test effects for significance, we ran mixed-effects models as implemented in the *lme4* package (Bates et al. 2015).

The acceptance rates per ratio and sentence type are presented in Table 2. There seems to be no difference between positive and comparative sentences in ratio 12:12. The difference is present for other ratios and forms the following pattern: 11:12< 9:12 = 7:12. We modeled the acceptance likelihood using a logit mixed-effects model, which included the main effects of sentence type and of ratio as predictors (fixed effects) and random intercepts for subject and for depicted object (random effects). There was a significant main effect of sentence type ($\beta = -3.06$, $SE = 0.15$, $z = -20.82$, $p < 0.001$). We then modeled the ratio effect on positive and comparative sentences separately (still including random intercepts for subject and depicted object). In sentences with a positive form of the adjective, ratio 9:12 did not

Table 2 Mean acceptance rates per ratio and sentence type, Experiment 2

Ratio	Sentence type	Total N trials	% of YES responses
7:12	Positive	365	54
	Comparative	370	96
9:12	Positive	366	54
	Comparative	369	96
11:12	Positive	365	39
	Comparative	369	93
12:12	Positive	362	11
	Comparative	356	8
Fillers	Positive	245	1
	Comparative	245	2

differ from ratio 7:12 ($\beta = 0.002$, $SE = 0.21$, $z = 0.01$, $p = 0.99$); 11:12 did differ from 7:12 ($\beta = -1.20$, $SE = 0.21$, $z = -5.66$, $p < 0.001$); and 12:12 also differed ($\beta = -3.60$, $SE = 0.27$, $z = -13.36$, $p < 0.001$). So the pattern 11:12< 9:12 = 7:12 for positive sentences is supported by the significance levels. In the model that included comparative sentences only, only ratio 12:12 significantly differed from 7:12 ($\beta = -7.18$, $SE = 0.47$, $z = -15.28$, $p < 0.001$).

Next, we looked at the differences between ratios and sentence types in terms of RTs. The mean RTs in 'yes' responses for each ratio are presented in Fig. 4. There seems to be a difference in RTs between sentence types in all ratios except for 12:12. In all cases, evaluating sentences with positive adjectives took longer than sentences with comparative adjectives. Moreover, the RTs increase with decrease in ratio sizes. In order to test for the statistical significance of these differences, we modeled RTs using a linear mixed-effects model. Here too, we included the main effects of ratio and sentence type as well as random intercepts for subject and depicted object. The p-values were calculated based on the Satterthwaite approximation for denominator degrees of freedom as implemented in the package *lmerTest* (Kuznetsova et al. 2016). There was a main effect of sentence type ($\beta = 155$, $SE = 15.85$, $t = 9.83$, $p < 0.001$). The RTs in ratio 9:12 did not significantly differ from ratio 7:12 ($\beta = 21.02$, $SE = 17.14$, $t = 1.23$, $p = 0.22$), whereas 11:12 did differ from 7:12 ($\beta = 120.5$, $SE = 17.8$, $t = 6.77$, $p < 0.001$) and 12:12 also did ($\beta = 297.52$, $SE = 37.7$, $t = 7.89$, $p < 0.001$). Thus, the pattern 11:12< 9:12 = 7:12 that we saw with acceptance rates is also present in RTs, but is valid for both positive and comparative sentences.

In order to examine whether all participants showed the same pattern of responses, we looked at inter-individual differences in acceptance rates. We calculated the acceptance rates for each sentence type in ratios 7:12, 9:12 and 11:12 combined. Ratio 12:12 was excluded since it is not of interest for our hypotheses. The resulting acceptance rates for each participant are presented in Fig. 5. We see that while some of the participants exhibit the effect that we expected and which we can see in the overall acceptance rates, other participants always responded 'yes' in both types of sen-

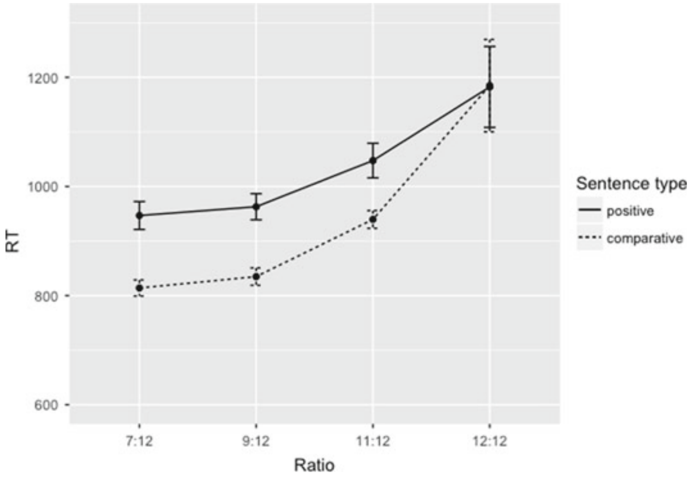


Fig. 4 Mean RTs per ratio and sentence type for 'yes' responses, Experiment 2. The error bars represent the standard error value

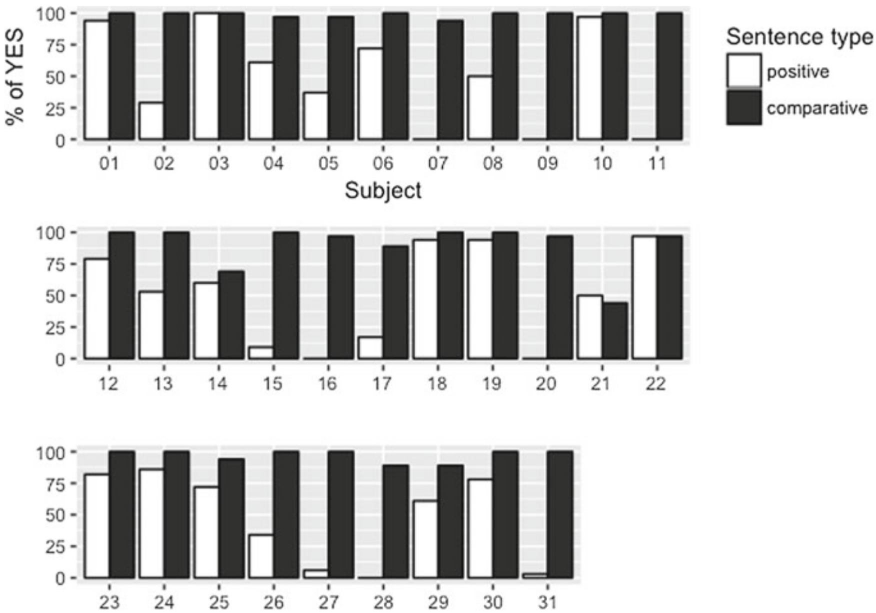


Fig. 5 Mean acceptance rates for ratios 7:12, 9:12 and 11:12 combined per sentence type, for each participant

Table 3 Mean acceptance rates per ratio and sentence type, *sensitive* group of subjects of Experiment 2

Ratio	Sentence type	Total N trials	% of YES responses
7:12	Positive	165	67
	Comparative	168	98
9:12	Positive	166	67
	Comparative	166	99
11:12	Positive	162	40
	Comparative	166	96
12:12	Positive	163	15
	Comparative	163	8
Fillers	Positive	112	2
	Comparative	111	1

tences, and others still never accepted a positive form and almost always accepted a comparative. In order to make sure that the effects that we obtained were not simply a result of averaging, we did the analyses again excluding all subjects for whom the difference between acceptance of positive and comparative forms was more than 90% and less than 10%. This resulted in an exclusion of 17 participants, or more than half. We call the remaining 14 subjects *sensitive*.

The acceptance rates for these 14 *sensitive* subjects are presented in Table 3. The pattern is the same as the overall average. The mean RTs are presented in Fig. 6. Here too, the pattern is the same as for the average of all subjects. We re-ran both models, looking at the acceptance rates and RTs, with these 14 subjects only. The pattern of the significant effects was the same.

Finally, we looked at the differences in acceptance rates between adjectives. We included all subjects in this analysis. We calculated the mean acceptance rate in all ratios except for 12:12. The resulting means are presented in Table 4. All adjectives seem to show the same pattern of acceptance rates.

Discussion As opposed to Experiment 1, here participants judged the applicability of positive forms of the vague adjectives and comparative forms differently. It appears that our suspicion that subjects interpreted positive forms as superlatives in Experiment 1 was correct.

We predicted that we would see the following acceptance rates for different ratios: 11:12< 9:12< 7:12; we also predicted the following RTs: 11:12>9:12>7:12. This prediction was not fully confirmed, but we did find partial support for our hypotheses: while 7:12 and 9:12 did not significantly differ from each other, they did differ from 11:12. A small gap in the case of ratio 11:12 was indeed less acceptable than the large gaps present in 9:12 and 7:12. One might have expected that in the ratio 11:12, the acceptance rate of the positive forms would be close to 0 for the *gap hypothesis* to be supported. We did not observe this. However, given that a considerable variability between subjects, and potentially between different depicted objects, is to be

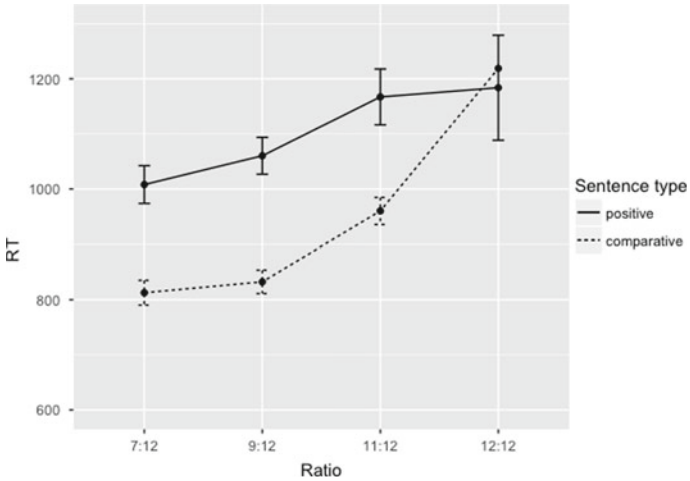


Fig. 6 Mean RTs per ratio and sentence type for ‘yes’ responses, *sensitive* group of subjects of Experiment 2. The error bars represent the standard error value

Table 4 Acceptance rates per adjective and sentence type, excluding ratio 12:12. Experiment 2

Adjective	Sentence type	Total N trials	% of YES responses
Big	Positive	366	44
	Comparative	370	95
Long	Positive	276	51
	Comparative	277	95
Tall	Positive	364	53
	Comparative	369	95
High	Positive	90	49
	Comparative	92	96

expected, we see a lower acceptance rate in ratio 11:12 than in larger ratios as strong enough evidence.

Related to the previous point, in the experiment planning stages, we had tentatively expected to observe the same acceptance rate for both positive and comparative forms in ratio 7:12, since the differences between the target and comparison objects was relatively large there. However, the mean acceptance rates for the positive form were never higher than 54% (or 67% in the *sensitive* group). One possible explanation for this is that perhaps the difference was still not large enough to satisfy our participants’ criteria. However, although the differences are not statistically significant, they do go in the expected direction in the case of the RTs: participants were on average faster at giving ‘yes’ responses in the ratio 7:12 than in the ratio 9:12. This difference was small overall, but 52 ms. for the *sensitive* group of subjects. A more careful inspection with more experimental trials and participants would be needed to make a definite conclusion about the RT differences there and why we did not see difference between these ratios in acceptance rates.

We obtained the predicted difference in RTs between the judgments of positive and comparative forms in our data. Giving a 'yes' response in the positive adjective judgment took longer than in the comparative form judgment. We take this as evidence that more intensive processing takes place when determining whether a positive form can indeed be applied. The additional process that needs to take place, we assume, is verifying that some criterion is met. Potentially, this criterion might be whether the difference is large enough.

Another aspect that came as a surprise and deserves closer inspection is inter-individual differences between our subjects. Only 14 out of 31 subjects that we tested exhibited the pattern of effects that we expected. For some of the subjects, there was no difference between comparative and positive forms, i.e. they always accepted both. We cannot say that these subjects simply pressed 'yes' in every trial, because they all gave correct 'no' responses to filler trials that we included. Thus, it seems like they interpreted the positive form as meaning something like 'one of the bigger ones' and therefore it was always true. Another group of participants never agreed that a vague adjective can be used as a valid description of our target object. They did always see a difference, because they agreed with the applicability of the comparative forms, so it cannot be that they simply did not notice size differences. Rather, it seems like they had some criterion that was never met in these experimental conditions. Potentially, they may have thought that a positive form can *only* be applicable when the object is the biggest/tallest etc., which was never the case in these pictures. However, this is just our speculation at this point. It would need to be tested in future studies.

It is also worth noting that the differences between both acceptance rates and the RTs for both positive forms and comparative forms are more pronounced in the group of *sensitive* subjects. The difference between RTs to positive and comparative sentences was around 200 ms for this group, whereas it was around 150 ms overall. These subjects seemed to have put extra effort into making a decision about the applicability of the positive form of the adjectives and that is why the expected difference is visible here.

2.3 Interpretation of the Experimental Data

We conducted two experiments that were the same except for one aspect and obtained very different results. Since the only difference between Experiment 2 and Experiment 1 was that we added a new object to the comparison objects in the pictures, we are inclined to conclude that in Experiment 1, participants thought it was acceptable to use a positive form because it referred to the biggest/tallest/longest/highest object on the screen (of course, for a definite conclusion an experiment directly comparing two situations would need to be conducted). In addition, there might have been a group of participants in Experiment 2 that had a strict requirement that the positive form can *only* be applied to the biggest/tallest, etc., objects. Overall, our evidence so far thus suggests that, at least in the set-up we used, the positive form of a vague adjective can be used as a description of the biggest/tallest etc. In other words, a gap between the target object and other objects is not necessary for the validity of a vague adjective if the target object is already the biggest.

In Experiment 2, we obtained evidence that there is a difference in the meaning of comparative and positive forms of the vague adjective. The comparative form was accepted in all ratios except for 12:12, so whenever there was any difference at all between the target and the comparison objects. This supports our *comparative hypothesis*. Moreover, we saw that judging the applicability of the positive form took longer than judging the applicability of the comparative form, so there is a difference also in the processing that took place in order to judge the applicability of two forms of the adjective.

The mean acceptance rate for the positive form of the vague adjective for the ratio 11:12 was lower than for the ratios 9:12 and 7:12. A parallel pattern of differences was obtained with reaction times. This supports the prediction of the *gap hypothesis*, i.e. the difference between the target object and other objects has to be significant in order for the positive form to be justified. At the same time, this pattern supports the *Weber's law hypothesis*: that with increasing relative differences in size/length etc., the applicability of the positive form of the vague adjective will increase.

3 Modeling by Semi-orders and Challenges

In Sect. 1, we formulated three hypotheses, which were put to the test using the experiments described in the previous section. The results that we obtained were consistent with the first and third hypotheses, but they were unexpected under the second hypothesis. In this section we will discuss whether we can account for the data making use of semi-orders.

3.1 Semi-orders and Indirect Versus Direct Comparisons

In van Rooij (2011a), it is proposed that the difference between explicit and implicit comparatives should be captured in terms of the difference between weak orders and semi-orders.[6]

A structure $\langle X, >_P \rangle$, with $>_P$ a binary relation on X, is a strict weak order just in case $>_P$ is irreflexive (IR), transitive (TR), and almost connected (AC):

Definition 1 A <u>strict weak order</u> is a structure $\langle X, >_P \rangle$, with $>_P$ a binary relation on X that satisfies the following conditions:

[6]See Solt (2016) for a recent use of semi-orders to account for the difference between the meanings of 'most' and 'more than half'.

(IR) $\forall x : \neg(x >_P x)$.
(TR) $\forall x, y, z : (x >_P y \wedge y >_P z) \rightarrow x >_P z$.
(AC) $\forall x, y, z : x >_P y \rightarrow (x >_P z \vee z >_P y)$.

It is well-known (e.g. Krantz et al. 1971) that strict weak orders can be represented numerically by a real valued measurement function f_P as suggested above: for all $x, y \in X$: $x >_P y$ iff $f_P(x) > f_P(y)$.[7] If we now define the indifference relation, '$\approx_P$', as follows: $x \approx_P y$ iff$_{def}$ neither $x >_P y$ nor $y >_P x$, it is clear that $x \approx_P y$ iff $f_P(x) = f_P(y)$. It follows immediately that the similarity relation is not only reflexive and symmetric, but also transitive: if $x \approx_P y$ and $y \approx_P z$, then $x \approx_P z$ holds as well. Thus, the indifference relation defined for (strict) weak orders is an equivalence relation.

In the linguistic literature (e.g. Kennedy 1999), comparisons based on adjectives like 'hot', 'tall', and 'loud' are normally based on exact measurement functions like our f_P above, which give rise to weak orderings. However, there are several phenomena for which a weak ordering seems already too strong. First, this is due to the phenomenon of *multi-dimensional adjectives*, which, arguably, give rise only to a (strict) *partial* order, i.e. an order that is just irreflexive and transitive. Second, and more relevantly for this paper, a weak order is too strong for phenomena related to vagueness. Because a weak order gives rise to an indifference relation that is transitive, we get into a Sorites problem:

> A person may be indifferent between 100 and 101 grains of sugar in his coffee, indifferent between 101 and 102, ..., and indifferent between 4999 and 5000. If indifference were transitive he would be indifferent between 100 and 5000 grains, and this is probably false. (Luce 1956).

To represent orders involving vagueness, the comparative '$x \succ_P y$' should intuitively mean that x is *significantly* or *noticeably* greater than y. Luce (1956) introduced orderings that capture exactly this, *semi-orders* [8]. A structure $\langle X, \succ_P \rangle$, with $\succ_P$ a binary relation on X, is a semi-order just in case $\succ_P$ is irreflexive (IR), satisfies the interval-order (IO) condition, and is semitransitive (STr).

Definition 2 A <u>semi-order</u> is a structure $\langle X, \succ_P \rangle$, with $\succ_P$ a binary relation on X that satisfies the following conditions:
(IR) $\forall x : \neg(x \succ_P x)$.
(IO) $\forall x, y, v, w : (x \succ_P y \wedge v \succ_P w) \rightarrow (x \succ_P w \vee v \succ_P y)$.
(STr) $\forall x, y, z, v : (x \succ_P y \wedge y \succ_P z) \rightarrow (x \succ_P v \vee v \succ_P z)$.

Although semi-orders cannot be adequately represented by measure functions that assign to each member of X just a single number, they can be represented by *intervals*. Scott and Suppes (1958) prove the following representation theorem for semi-orders:

[7] This is true in general only for *countable* sets. Things are more complicated otherwise.
[8] See also Rubinstein (1988).

Fact 1 *Let X be a finite set and $\epsilon > 0$. Then $\langle X, \succ_P \rangle$ is a semi-order iff there exists a real-valued function g_P on X such that for all $x, y \in X$: $x \succ_P y$ iff $g_P(x) > g_P(y) + \epsilon$.*

To illustrate, suppose we have a semi-order $\langle X, \succ_P \rangle$ with $x \sim_P y \sim_P z \sim_P v$ but $x \succ_P z$, $y \succ_P v$ and $x \succ_P v$. If we assume that $\epsilon = 1$, this can be modelled by the following function g: $g_P(x) = 4$, $g_P(y) = 3.2$, $g_P(z) = 2.4$, $g_P(v) = 1.6$.

As for weak orders, for semi-orders too we can define a similarity relation $\sim_P$ in terms of '$\succ_P$' as follows: $x \sim_P y$ iff neither $x \succ_P y$ nor $y \succ_P x$. The relation '$\sim_P$' is reflexive and symmetric, but need not be transitive. Thus, '$\sim_P$' does not give rise to an equivalence relation. And this is required to solve the Sorites problem: small differences can be neglected, but many small differences together add up to significant differences, and these cannot be neglected. Measure theoretically '$x \sim_P y$' is true iff the difference in height between x and y is less than ϵ. In case $\epsilon = 0$, the semi-order is a weak order.

As already suggested above, weak orders are very natural representations of standard explicit comparatives like (1a). In van Rooij (2011a) it was proposed that the semi-order relation *significantly taller than*, i.e. '$\succ_T$', is what is relevant to evaluate the truth of implicit comparatives like (1b). Thus, (1b) is true just in case John is significantly taller than Mary. This immediately explains why (1a) can be inferred from (1b), but not the other way around.

3.2 Challenges for the Semi-Order Approach

The pragmatic gap requirement The *gap hypothesis* seems confirmed by the results of Experiment 2 as described above. Given that with a semi ordering we can think of $x \succ_{Tall} y$ as true just in case x is *significantly* taller than y, it seems we can naturally account for the pragmatic gap requirement involving clustering a group into the tall ones versus the rest.

However, the results of Experiment 1 suggest that the hypothesis that a significant gap is a necessary condition for dividing groups is too strong[9]: even if there is no clear demarcation between the bigger and the smaller persons of the domain, the tallest person can still be called 'tall'.[10] Thus, the *gap hypothesis* seems not to allow for such exceptional situations, and the existence of a significant gap does not seem necessary. But once we are in this situation one wonders how to address the Sorites paradox.

Semi-orders and Weber's law Another challenge is how the semi-order approach can account for Weber's law. In the introduction we argued that under time-pressure

[9]There is, of course, another problem as well: it seems that the existence of a significant gap is not *sufficient* either. Think of the case where an ordered set of objects contains more than one gap with respect to height. Which gap should be used to distinguish the tall ones from the rest?

[10]Since it was not an a-priori hypothesis that tested here, but a conclusion we came to based on looking at the data, this claim will need to be explicitly tested in a separate experiment in the future.

people represent magnitudes at least also in terms of so-called Analogue Magnitude Representations (AMRs). Fults (2011) has argued that this is relevant for how to interpret natural language comparatives like (1a) and (1b) and positives like (2). He claims that as a result of our use of AMRs for the analysis of such sentences, we can explain why there are no sharp cut-off points for the application of a predicate like 'tall' and that its application gives rise to variable uncertainty. Moreover, it explains immediately why vague predicates obey Weber's law, and why what is a sufficient gap between two magnitudes is determined by their ratio and grows with intensity.

In van Rooij (2011a), semi-orders were argued to be important for the analysis of implicit comparatives and of positive sentences involving vague predicates. Based on the abobe hypothesis, however, Fults (2011) comes up with a number of challenges for the approach based on semi-orders whose numerical representation (via Fact 1) is based on a precise and fixed threshold ϵ.[11]

1. ϵ is too sharp a cutoff point: analogue magnitudes and vagueness do not have sharp points of discrimination (the problem of higher-order vagueness).
2. semi-order are incapable of providing an account of variable uncertainty, since again, two values $f_P(x)$, $f_P(y)$ are either within ϵ to each other or not; there are no intermediate levels.
3. ϵ is fixed, while analogue magnitudes and vagueness show scalar variabilities. If semi-orders are to be used for either, ϵ must change based on the value of the magnitudes being compared.

We would like to suggest that these three critiques involve two different intuitions: the first two critiques involve the *sharpness* of the threshold ϵ, while the third critique involves, rather, the *fixed* character of ϵ. We will respond to these critiques by claiming that both intuitions are correct, but that they can still be captured in terms of semi-orders. In Sect. 4.2 we will show that for orderings to be semi-orders, cutoff points do not have to be sharp, but can also be *probabilistic* in nature. This will deal with the first two critiques. In the same section we will show that orderings where the threshold increases with increasing magnitude can be modelled in terms of semi-orders as well.

A final problem concerns the *interpretation* of the threshold. The representation theorem for semi-orders makes use of a threshold, ϵ, which we interpreted as a *significant gap*. The most natural interpretation of this threshold, however, is as a *just noticeable difference*, abbreviated by JND, or *discrimination threshold*. It is also this interpretation of the threshold which explains why magnitudes observe Weber's law. But one wonders whether the notions 'significant gap' and JND can be identified. According to Kennedy's (2011) intuition, for instance, an explicit comparative can be true, though the corresponding implicit comparison is false. But if semi-orders with a threshold ϵ are used to account for the falsity of the implicit comparison, it is

[11] According to proponents of AMRs, these representations are associated with arithmetic computations, including comparison, addition, subtraction, multiplication, and division. If so, one can define another challenge for our representation based on semi-orders: how to account for this? We will ignore this challenge in this paper, but see Solt (2016) for a suggestion of how to meet this challenge.

hard to explain why the explicit comparison can still be true, when ϵ is interpreted as the JND. Is there any relation between our interpretation of ϵ and the standard one as a JND?

4 Answering the Challenges

4.1 Revisiting the Gap Hypothesis

The results of Experiment 2 show that the *gap hypothesis* is quite natural. The results of Experiment 1 suggest, however, that the existence of a gap is not a necessary condition for the appropriate use of sentences like 'John is tall': also in case there is no gap, the largest individual is still counted as tall. Of course, one could conclude from Experiment 1 that the *gap hypothesis* should thus be rejected completely. However, we still feel that the existence of gap for the appropriate use of predicates is still a natural *default*. How can we account for such a default?

In Cobreros et al. (in press) a non-monotonic consequence relation $\models^{PrPr}$ is introduced to account for the intuition that although normally we conclude $Tall(y)$ from premisses $Tall(x)$ and $x \sim_{Tall} y$, this conclusion can be cancelled in case additional information is provided (in this case that x is a borderline case of a tall person). According to this logic, it follows by default that in case x is considered tall and y is not, y is not considered to be similarly tall as x, i.e., $x \not\sim_{Tall} y$. Interpreting the similarity relation as the indifference relation of a semi-order, this means that there is a significant gap in terms of height between x and y. More generally, if we have a sequence of individuals $x_1, \ldots x_n$, the first of which is obviously tall and the last obviously not, this analysis predicts that, by default, somewhere in the sequence there is a gap, i.e., a pair x_i, x_j such that $x_i \not\sim_P x_j$. We take this to be in accordance with the *gap hypothesis*. However, just as before, this default inference can be cancelled. Also when it is explicitly given as premisses that the first individual in the sequence is tall and the last one is not, it is still allowed by the logic that $x_1 \sim_{Tall} x_2 \wedge \cdots \wedge x_{n-1} \sim_{Tall} x_n$. As such, the Sorites problem is (dis)solved. Moreover, it is possible in this logic that there exists an i, $1 \leq i \leq n$, such that $\forall j : i < j \leq n : x_j$ is not counted as tall.[12] It could be, for instance, that this ith individual is x_2, as observed in Experiment 1. Thus, this logic not only captures the idea that the use of a predicate like 'tall' normally comes with a gap, it is also *consistent* with the results of Experiment 1 that only the tallest individual in the sequence is classified as tall. This logic does *not predict* the results of Experiment 1, however, for the logic also allows any other individual in the sequence to be the last tall individual. At this point we don't know yet whether this means that the logic $\models^{PrPr}$ is thus too liberal, or

[12]In the logic $\models^{PrPr}$, be borderline tall is modelled by saying that you are both tall and not tall. Although Cobreros et al. (in press) claim that there is good empirical motivation for this modeling of borderline cases, this particular modeling of borderline cases is not required for a non-monotone logic like $\models^{PrPr}$ to work.

whether people find it indeed quite arbitrary which individual is the last tall person. For now, we leave this as it is, but only conclude that the results of Experiment 1 are not in contradiction with the *gap hypothesis*, at least once one weakens this principle—expressible as $(Tall(x) \wedge \neg Tall(y)) \rightarrow x \not\sim_{Tall} y$—to a default.

4.2 Meeting Fults' Objections

Cutoff-points and uncertainty On the intuitive understanding of the numerical representation of semi-orders, the threshold ϵ represents *uncertainty* as to whether x is A-er than y or not. What the standard numerical representation does not capture, however, is that this uncertainty is, in fact, not equally divided over ϵ, when the latter is thought of as an interval. What one intuitively wants to capture to account for the first two critiques of Fults (2011) is that the uncertainty peaks at the centre of the interval, and fades away at the edges. Uncertainty is naturally represented in terms of probability, and we propose to answer Fults' first two critiques in terms of *probabilistic choice* models.

There are typically two interpretations for axioms that define orders, such as asymmetry and almost connectedness. On a *normative approach*, these axioms define what it means to be rational: an individual who violates the maxims is acting irrationally. This is typically assumed in decision theory, where an ordering is used to model preference. On a *descriptive approach*, on the other hand, the axioms describe testable conditions. Taking a descriptive approach, we see that individuals, or groups, often show inconsistent behavior. For instance, a subject might at one time find that sound x is louder than sound y, while at other times (s)he might make the opposite judgement. Or it might be that some agents of a group judge sound x as louder than sound y, while other agents make the reverse judgement. Still, even in these cases the agent, or group, might behave *probabilistically consistent*. Let P^A_{xy} be the frequency with which an agent considers x A-er than y for some adjective A, or the proportion of individuals that considers x A-er than y. If it is assumed that $P^A_{xy} + P^A_{yx} = 1$ for all $x, y \in X$, $\langle X, P^A \rangle$ is called a *forced choice pair comparison system* in mathematical psychology. To make sense of the notion of probabilistic consistency as mentioned above, a measurement representation can be given to such a system. On one natural proposal, what P^A_{xy} measures is a comparative strength of A-ness, and this can be captured by a real-valued function f_A as follows: $P^A_{xy} > P^A_{yx}$ iff $f_A(x) > f_A(y)$ (cf. Luce and Suppes 1965). Under some natural conditions, this kind of ordering relation will give rise to a (strict) weak order. But, of course, many times we would say that an agent or group only considers x A-er than y if x is judged to be A-er than y a sufficiently large percentage of the time, or by a sufficiently large proportion. In such a case we can define a relation $\succ^{\epsilon}_A$ on X as follows: $x \succ^{\epsilon}_A y$ iff $P^A_{xy} > \epsilon$, with $\epsilon \in [\frac{1}{2}, 1]$. Luce (1958) has already shown that this ordering relation is a semi-order, in case the forced choice comparison system $\langle X, P^A \rangle$ satisfies certain 'natural' conditions.[13] Of course, if $\epsilon = \frac{1}{2}$, the semi-order is a strict weak order.

How is this all related to Fults' first two critiques? That is, how to make use of forced-choice comparison systems to capture the intuition that the uncertainty peaks at the centre of the interval ϵ and fades away at the edges? To answer that question, notice that in case $P^A_{xy} = \frac{1}{2}$ there is maximal uncertainty about whether x is A-er than y or not. In case $P^A_{xy} = 1$, x is considered to be certainly A-er than y. Fults' (2011) first two critiques are thus captured once we make the natural assumption that $P^A_{xy} = \frac{1}{2}$ exactly at the centre of the interval ϵ, marking the difference between x and y and that P^A_{xy} approaches 1 or 0 at the edges.

Weber's law and semi-orders Revealing as the representation theorem of Scott and Suppes for semi-orders (Fact 1) may be, it seems to contradict Weber's law, according to which the threshold grows with intensity. But if this threshold grows with intensity, and thus is not constant, to capture Weber's law algebraically the representation theorem involving semi-orders (Fact 1) does not seem to be appropriate. What we need is a representation theorem where the threshold is *variable*.

As it happens, there exists a representation theorem with variable thresholds for so-called interval orders, that was introduced by Wiener (1914). The structure $\langle X, \rhd \rangle$ is an interval order in case the relation $\rhd$ is irreflexive, and satisfies the interval-order condition:

Definition 3 An interval order is a structure $\langle X, \rhd \rangle$, with $\rhd$ a binary relation on X that satisfies the following conditions:
(IR) $\forall x : \neg(x \rhd x)$.
(IO) $\forall x, y, v, w : (x \rhd y \wedge v \rhd w) \rightarrow (x \rhd w \vee v \rhd y)$.

Notice that any interval order is also a strict partial order, because (IR) and (IO) together entail that the resulting ordering is transitive. Just like for strict partial orders and semi-orders, the indifference relation ($x I_{\rhd} y$ iff $\neg(x \rhd y) \wedge \neg(y \rhd x)$) for interval orders need not be transitive, and thus does not give rise to an equivalence relation. As a result there is no possibility of getting a full numerical representation, where each member of X is mapped to a single number. But just like semi-orders, interval orders can also be represented in terms of threshold (Fishburn 1970), but this time this threshold need not be fixed.

Fact 2 *Let X be a finite set and $\epsilon > 0$. Then $\langle X, \rhd \rangle$ is an interval-order iff there exists two positive real-valued functions g and ϵ such that for all $x, y \in X$: $x \rhd y$ iff $g(x) > g(y) + \epsilon(y)$, with $\epsilon(y) \geq 0$ for all y*

We know that every semi-order is an interval order. Indeed, this just follows if we take ϵ to be a constant function. However, it also follows that the semi-order discussed in Sect. 3.1 just below the statement of Fact 1, $\langle X, \succ_P \rangle$ with $x \sim_P y \sim_P$

[13] In particular, when $\langle X, P^A \rangle$ satisfies *strong stochastic transitivity*: $\forall x, y, z \in X$: if $P^A_{xy} \geq \frac{1}{2}$ and $P^A_{yz} \geq \frac{1}{2}$, then $P^A_{xz} \geq max\{P^A_{xy}, P^A_{yz}\}$.

$z \sim_P v$ but $x \succ_P z$, $y \succ_P v$ and $x \succ_P v$, can be represented *as an interval order* by the following functions g and ϵ: $g(x) = 2$, $g(y) = 1.5$, $g(z) = 1.1$, $g(v) = 0.8$, and $\epsilon(x) = 0.7$, $\epsilon(y) = 0.6$, $\epsilon(z) = 0.5$ and $\epsilon(v) = 0.4$, if we define $x \rhd y$ iff $g(x) > g(y) + \epsilon(y)$. Still, not all interval orders are semi-orders: it is possible that we have an interval order $\langle X, \rhd \rangle$ such that $x \rhd y \rhd z$, but where it holds that $vI_{\rhd}x$, $vI_{\rhd}y$ and $vI_{\rhd}z$. This is possible in case $\epsilon(v)$ is very large compared to $\epsilon(x)$, $\epsilon(y)$ and $\epsilon(z)$.

Fact 2 shows that the phenomena that are captured by Weber's law can be represented by an interval-order. However, we will show that although Weber's law has variable thresholds, it can still be represented in a qualitative way as a semi-order.

It is interesting to observe what happens if we demand that for $x \rhd y$ to hold, we demand not only that $g(x) > g(y) + \epsilon(y)$, with $\epsilon(y) > 0$ for all y, but also make the following constraint: $g(x) \gtrless g(y)$ iff $\epsilon(x) \gtrless \epsilon(y)$. Notice that it now follows that xIy iff $g(x) \not> g(y) + \epsilon(y)$ and $(g(y) \not> g(x) + \epsilon(x)$. But this means that xIy holds iff $|g(x) - g(y)| \leq min\{\epsilon(x), \epsilon(y)\}$. With this in hand, we can show that our example of an interval-order that is not a semi-order cannot be represented by the functions g and ϵ observing the new constraint: it is impossible that $xPyPz$, but where it holds that vIx, vIy and vIz. For if vIy, then either (i) $min\{\epsilon(v), \epsilon(y)\} = \epsilon(v)$ or (ii) $min\{\epsilon(v), \epsilon(y)\} = \epsilon(y)$. But if (i), then $g(v) \leq g(y)$, and thus $x \rhd v$, which contradicts our assumption that xIv. And if (ii) then $g(v) \geq g(y)$, from which it follows that $v \rhd z$, which contradicts our assumption that $vI_{\rhd}z$.

The idea that the threshold grows with intensity means that for all $x, y \in X$: $g(x) \gtrless g(y)$ iff $\epsilon(x) \gtrless \epsilon(y)$. In Appendix B, we show that any interval-order that is represented by functions g and ϵ satisfying this constraint is, in fact, a semi-order.

Fact 3 *If we can define an order $\langle X, P \rangle$ in terms of two real-valued measure functions g and ϵ on X such that for all x, y: xPy iff $g(x) > g(y) + \epsilon(y)$, with $\epsilon(y) > 0$ for all y and for all x and y: $g(x) \gtrless g(y)$ iff $\epsilon(x) \gtrless \epsilon(y)$, then this order is a semi-order.*

Observe that the 'if' cannot be turned into a 'if and only if'. The above numerical representation is not a necessary condition for semi-orders. Let us provide two examples to show this. First, Fact 1 shows that in order for xPy to hold in a semi-order, it doesn't have to hold that $\epsilon(x) > \epsilon(y)$, if $g(x) > g(y)$: a constant ϵ will do. Second, recall that any strict weak order is also a semi-order. However, there are strict weak orders that cannot be numerically represented by the above, because almost connectedness (i.e., $\forall x, y, z(xPy \rightarrow (xPz \vee zPy)))$ does not hold, as can be shown by the following counterexample: $g(x) = 1$, $g(z) = 0.9$ and $g(y) = 0.8$ and $\epsilon(x) = 0.25$, $\epsilon(z) = 0.2$, and $\epsilon(y) = 0.15$. Then it is clear that although xPy because $g(x) = 1 > 0.95 = g(y) + \epsilon(y)$, it is neither the case that xPz (because $g(x) = 1 \not> 0.9 + 0.2 = g(z) + \epsilon(z))$ nor that zPy (because $g(z) = 0.9 \not> 0.8 + 0.15 = g(y) + \epsilon(y))$. In other words, even though the above numerical representation of an order $\langle X, P \rangle$ is a *sufficient* condition for it to be a semi-order, it is clear that this numerical representation is *not* a *necessary* one.

Notice also that the relation I is not only reflexive, but because of xIy iff $|g(x) - g(y)| \leq min\{\epsilon(x), \epsilon(y)\}$ *also* symmetric. The similarity relation does *not* have to be transitive, as our earlier example illustrates: xIz and zIy, but xPy.

From fact 3 it immediately follows that Weber's law can be captured in terms of a semi-order:

Fact 4 *Weber's law can be captured in terms of a semi-order.*

Recall that Weber's law states that the size of the difference threshold, $\Delta g(x)$, is lawfully related to $g(x)$, the magnitude of stimulus x as follows: $\frac{\Delta g(x)}{g(x)} = k$. In this 'law', $\Delta g(x)$ represents the difference threshold, $g(x)$ represents the stimilus intensity of x, and k signifies that the proportion on the left side of the equation remains constant despite variation for the $g(x)$ term. Stating it otherwise, $\Delta g(x) = k \times g(x)$. If the difference threshold is a constant proportion equal to 0.1, then the size of the threshold for an experience caused by something with an intensity of 100 units would be 10 units (i.e., $\Delta g(x) = 0.1 \times 100 = 10$), and an experience caused by something with an intensity of 1000 units would be 100 units (i.e., $\Delta g(x) = 0.1 \times 1000 = 100$). Now we can say that xPy iff $g(x) > g(y) + \Delta g(y)$ and that $xI_P y$ iff $|g(x) - g(y)| \leq \Delta g(y)$. We have represented interval-orders above as follows: xPy iff $g(x) > g(y) + \epsilon(y)$. These two are obviously equivalent, if we take $\epsilon(y)$ to be $\Delta g(y)$ for each y. However, because it now immediately follows that $\epsilon(x) > \epsilon(y)$ iff $g(x) > g(y)$, our interval order is, in fact, a *semi-order*, because this was the condition we required for an interval order to be a semi-order.

4.3 The Interpretation of the Threshold

In the beginning of the paper we followed Kennedy's (2011) proposal that whereas for the truth of the *explicit* comparative (1a) (repeated as (3a)) any (directly or indirectly observable) bit of John's height that exceeds Mary's suffices, there should be a *significant gap* between the heights of John and Mary for the *implicit* comparative (1b) (repeated as (3b)) to be true.

(3) a. John is taller than Mary.
b. Compared to Mary, John is tall, but compared to John, Mary is not tall.

Later in the paper, however, we interpret a *significant gap* as the standard interpretation of a threshold (Luce 1956): as a *just noticeable difference*. But this suggests that whereas (3b) can only be true if there exists a JND between the heights of John and Mary, no such JND need to exist in order for (3a) to be true. But if no JND need to exist between the heights of John and Mary, one wonders how (3a) could be true at all? Perhaps one could say that for the truth of (3b) it is only demanded that there be a difference, not that it is noticeable. Perhaps. We take it to be more natural, however, that two notions of just noticeable difference are involved, although both depend on length.

Perhaps economics provides the most obvious case to illustrate two different ways to think about just noticeable differences that depend on the same objective measure. Bernoulli (1738) has proposed a difference between perceived happiness and amount of money, although the former might depend on the latter. Although it is clearly the case that after counting, a person can notice the difference between €995 and €1000 equally well as the difference between €15 and €20, it can still be the case that this person cannot perceive a difference in happiness between owing €995 and €1000, but can do so between owing €15 versus €20. Perceived amount of money clearly gives rise to a non-zero notion of just noticeable difference in both cases, while perceived pleasure need not be. This is why Bernoulli concluded that utility should not be measured in terms of euros (or florins, in his case). Although utility depends on euros, the dependence is logarithmic rather than linear in nature. One might think of Bernoulli's strategy for making a distinction between amount of euros and utility in a slightly different way: as one between *exact counting*, on the one hand, and *approximate counting*, on the other.[14] Thus, in both cases, it is only numbers of euros that counts, but in the one case this number is more exact than in the other case.

Can we explain the results of Experiment 2, and Kennedy's (2011) intuition, according to which we would not count John as being tall (compared to Mary), although he is still observed to be taller, in terms of a similar strategy? In van Rooij (2011a) it was proposed that the difference between explicit and implicit comparatives like (3a) and (3b), and the difference between weak orders and semi-orders, can be thought of in terms of the distinction between a direct versus indirect observable difference: whereas direct observable difference gives rise to a semi-order, indirect observable difference gives rise to a more fine-grained weak order. But what does it mean to be a direct or indirect observable difference? One way to account for the distinction between the two is to assume that what counts for 'direct distinguishability' is speed, and the ability to make the distinction under high cognitive load. It might well be that although precise counting and exact measuring is possible when enough time is allowed, this is not possible under these time-pressured circumstances. What counts in these latter cases is approximate magnitudes, and these give rise to Weber's law. Thus, we propose that we should interpret the thresholds involved in our use of semi-orders to account for the data as just noticeable differences, and to explain (Kennedy's 2011) intuition—with which we agree—that there is a difference (in English) between the use of explicit versus implicit comparatives in terms of the circumstances under which we would use such comparatives. In contrast to explicit comparatives, implicit comparatives are used only in case precise measures do not matter, or when there is not time enough to determine them. We propose a similar hypothesis to explain the data of Experiment 2: the use of 'tall', instead of the more precise 'taller than', is limited to those circumstances in which all that counts, or all

[14]But as one reviewer rightly remarked, this cannot be everything to Bernoulli's strategy: Bernoulli also introduced logarithmic functions for purely formal reasons, i.e., to prevent choices from having infinite expected utility. There are also general considerations about the diminishing marginal value of money.

that can be done, is (to make) a division of a group into a few relevant classes based on (approximate) height.[15,16]

5 Conclusion

The problem of how we group individuals or objects when we use a vague adjective like *tall* is one that has puzzled linguists and philosophers for a long time. Unlike the case of a comparative form of such adjectives, where it seems like the definition is straightforward, we do not know what exact criteria are used to determine when a positive form of a vague adjective is appropriate as a description. In this paper, we focused on the *gap hypothesis* as a potential solution. According to this proposal, for example in the case of height, we would only use *tall* if there was a significant gap between the shortest person in the group of people classified as *tall* and the tallest person in the *non-tall* group. This hypothesis is related to the fact that we as humans do not have precise information about, for example, heights of people. In a situation where we cannot take a ruler and measure the person's height, we make an estimate. For such a situation, Weber's law states that discriminability is determined by the ratio of difference between the two values and the higher the difference ratio, the easier it is for us to discriminate them.

We have presented experimental evidence that confirms this intuition about the difference in meaning between the positive and comparative forms of the vague adjectives: while any difference in sizes between the target object and the other objects within the context was sufficient for the comparative forms to be acceptable, acceptability of the positive form depended on the size difference. For using the positive form as a description, the larger size differences resulted in higher acceptance rates (and lower reaction times) than smaller ones. Our results also suggest that a positive form can in principle be used to describe simply the biggest, longest, etc., object within the context regardless of whether there is a gap or not.

One way to capture the difference between the comparative and positive forms existence of which our experiments confirmed is in terms of weak orders and semi-orders (van Rooij 2011a). However, a number of challenges to semi-orders have been raised by Fults (2011) based on Weber's law and Analoge Magnitude Representations. Namely, semi-orders make use of a sharp threshold of applicability that would not be possible if we indeed make approximations of values such as people's heights using AMRs. In addition, so far it has been proposed that the threshold is fixed across different values, which again is problematic under AMRs that are rather ratio-based. Besides these challenges, our experimental results showed that the requirement of

[15] Of course, one can think of circumstances where we talk about high magnitudes, but still a very fixed threshold is involved. For instance, it might be that one only has to pay a higher percentage of income taxes when one earns more than exactly €100,000.

[16] This hypothesis would also explain, we believe, why, intuitively, the threshold has to be larger for the case of positives like (2) than for (implicit) comparatives.

a significant gap might be too strong for vague adjectives. In Sect. 4, we showed how the analysis of vagueness presented in van Rooij (2011a) can be adjusted to meet these objections and refer to ongoing work that can solve the latter issue with the significant gap not being required. We thus showed that an analysis in terms of weak orders and semi-orders can still account for the use of vague adjectives and the difference between comparative and positive forms.

One question that we feel is still up for discussion is the interpretation of the threshold for comparative and positive forms that we discussed in Sect. 4.3. Intuitively, and based on our experimental results, an explicit comparative such as *taller* can be used for a situation where a positive form *tall* would not be acceptable. So there should be two different thresholds. We put forward our explanation of the difference as being in the contexts where the positive and comparative are used. We suggested that the positive form is used in situations when we cannot make exact measurements but just approximate ones, and that explains why Weber's law is important there, whereas a comparative form would be used when precise values are known and are important.

There should be follow-up experimental work to investigate the exact processes that result in differences in reaction times between the comparative and positive adjective judgments that we obtained. In addition, we need to find out where the differences between participants stem from. Our participants either interpreted the meaning of vague adjectives differently or interpreted the set-up differently. It would be useful to know whether their choices could be altered if they were given a different task. For example, they could be given a more explicit communicative task where they have to describe an object to another person.

Appendix A. Experimental Procedure and Participants

Procedure The experiment was built using a JavaScript library for online chronometric experiments—JsPsych—and run in participants' web browsers. This library allows for collection of participants' responses as well as fairly accurate reaction time data (Leeuw 2015; Leeuw and Motz 2016). The first screen that participants saw displayed general information about the experiment. On the second screen, they had to give consent to participate, and agree with storage of the data obtained. They then filled in a short questionnaire regarding their background, which served as a check that indeed all eligibility requirements were met. This was followed by detailed instructions for the task. Participants then did three practice trials, after which the experiment itself started. There were two blocks—one with sentences with positive vague adjectives and one with sentences with comparative vague adjectives. Each of the blocks contained 56 trials. Between the blocks, participants could take a break.

Trials within each block were presented in random order. Each trial started with a display of the sentence in the middle of the screen. When the participants finished reading the sentence, they pressed the space bar to see the picture itself. They then had to press either P if they agreed that the sentence could be used to describe this picture, or Q if they did not agree. The participants had 2300 ms to give a response, and if no response was given, the trial ended automatically and it was recorded as a missing response. The next trial then started after a random inter-stimulus interval between 700 and 1200 ms.

The instructions that participants received were the following:

> In this experiment, you will see pairs of sentences and pictures. Your task is to indicate whether you think the sentence is a good description of the given picture. In other words, do you think people would use this description for the given picture? You will need to think about the exact meanings of words in order to make correct judgments.
>
> Every trial will start with a sentence [...] Please pay attention to the meanings of sentences and give correct answers, but also try to do it quickly [...].

Participants Eligible participants had to meet the following criteria: native speaker of English; age 18–35 years old; born and currently living in the USA; right-handed. Completing the experiment should have taken approximately 15 min and the participants were paid 1.80 £.

Experiment 1: Twenty-five participants completed the task. One participant was excluded due to reading the instructions for under 10 s. The remaining 24 participants were included in the analysis. Nine were male, 15 female; their mean age was 24.96 (range 18–33). They took 11:29 min on average to complete the task (min. 08:33 and max. 17:11).

Experiment 2: Forty participants completed the task. One participants reported being color-blind and was excluded. Six further participants were excluded for reading the instructions for under 10 s. Two participants were excluded for pressing a single button as a response throughout the experiment. Thus, 9 participants in total were excluded and the analyses were performed on the remaining 31 participants. The mean age of these participants was 26.26 (range 20–34). Thirteen were female, 17 male, and 1 of other gender. Participants took 11:19 on average to complete the task (min. 07:11, max. 17:32).

Appendix B. Proof of Fact 3

To prove Fact 3, we have to show that the resulting order obeys the conditions for being a semi-order. That is, it has to be (i) irreflexive (IR), (ii) satisfy the interval order condition (IO), and (iii) be semi-transitive (STr).

(IR) Irreflexivity follows immediately if $\epsilon(x) \geq 0$.

(IO). Suppose xPy and vPw. To prove xPw or vPy. Suppose $\neg xPw$. It follows that (i) $g(x) > g(y) + \epsilon(y)$, and (ii) $g(v) > g(w) + \epsilon(w)$, and (iii) $g(x) \not> g(w) + \epsilon(w)$.

Now either (a) $\epsilon(v) = \epsilon(x)$, (b) $\epsilon(v) < \epsilon(x)$, or (c) $\epsilon(v) > \epsilon(x)$.

(a) But then by the constraint it follows that $g(v) = g(x)$, and with (ii) and (iii) we have a contradiction.

(b) But then by the constraint it follows that $g(x) > g(v)$ and thus with (ii) $g(x) > g(w) + \epsilon(w)$, which is in contradiction with (iii).
(c) But then by the constraint it follows that $g(v) > g(x)$ and thus with (i) $g(v) > g(y) + \epsilon(y)$. But then vPy.

Similarly, we can prove that if $\neg vPy$ it follows that xPw, which is enough to prove what we wanted.

(STr). Suppose $x > y$ and $y > z$. To prove for any v: xPv or vPz. So suppose xPy and yPz and $\neg xPv$. That is, suppose (i) $g(x) > g(y) + \epsilon(y)$ (ii) $g(y) > g(z) + \epsilon(z)$, and (iii) $g(x) \not> g(v) + \epsilon(v)$. Now either (a) $\epsilon(v) = \epsilon(x)$, (c) $\epsilon(v) > \epsilon(x)$ or (c) $\epsilon(v) < \epsilon(x)$.

(a) By the constraint it follows that $g(x) = g(v)$. Notice that from (i) $g(x) > g(y) + \epsilon(y)$ and (ii) $g(y) > g(z) + \epsilon(z)$ it immediately follows that $g(x) > g(z) + \epsilon(z)$ (transitivity). Thus vPz.
(b) By the constraint it follows that $g(v) > g(x)$. Notice that from (i) and (ii) it follows that $g(x) > g(z) + \epsilon(z)$, as in (a). Thus now also $g(v) > g(z) + \epsilon(z)$. But this means that vPz.
(c) By the constraint it follows that $g(x) > g(v)$, and thus $g(v) + \epsilon(v) \geq g(x) > g(v)$. Now there are three possibilities: either (c1) $\epsilon(v) = \epsilon(y)$, (c2) $\epsilon(v) < \epsilon(y)$, or (c3) $\epsilon(v) > \epsilon(y)$.
    (c1) By the constraint it follows that $g(y) = g(v)$. But this is impossible, because we have assumed that xPy and $\neg xPv$.
    (c2) By the constraint it follows that $g(y) > g(v)$. Thus, $g(y) + \epsilon(y) > g(v) + \epsilon(v)$. But this is impossible, because we have assumed that xPy and $\neg xPv$.
    (c3) So, (c3) has to be the case. But this means with the constraint that $g(v) > g(y)$. Because $g(y) > g(z) + \epsilon(z)$, it follows that $g(v) > g(z) + \epsilon(z)$. Thus vPz.

Thus, if $\neg xPv$, then vPz.

Similarly, we can also prove that if $\neg vPz$, then xPv. Now we have proved semi-transitivity.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii academiae scientiarum imperialis Petropolitanae*, *5*, 175–192.

Cobreros, P., Egre, P., Ripley, D. & van Rooij, R. (in press). Tolerant reasoning: nontransitive or nonmotonic? *Synthese*.

Dehaene, S. (2011). *The number sense: how the mind creates mathematics*. Oxford: Oxford University Press.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12.

Égré, P. (2017). Vague judgment: a probabilistic account. *Synthese*, *194*(10), 3837–3865.

Evans, J. S. B. T. (2010). *Thinking twice: two minds in one brain*. Oxford: Oxford University Press.

Fara, D. G. (2000). Shifting sands: an interest-relative theory of vagueness. *Philosophical Topics*, *28*, 45–81.

Fishburn, P. C. (1970). Intransitive indifference with unequal indifference intervals. *Journal of Mathematical Psychology*, *7*(1), 144–149.

Fults, S. (2011). Vagueness and scales. In P. Egre & N. Klinedinst (Eds.), *Vagueness and language use* (pp. 25–50). New York: Palgrave Macmillan.

Gaifman, H. (2010). Vagueness, tolerance and contextual logic. *Synthese*, *174*(1), 5–46.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Garland Press.

Kennedy, C. (2011). Vagueness and Comparison. In P. Egre & N. Klinedinst (Eds.), *Vagueness and language use* (pp. 73–97). Basingstoke: Palgrave Macmillan.

Krantz, D., Luce, D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement I: Additive and polynomial representations* (Vol. 1). New York: Academic.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models.

Luce, D., & Suppes, P. (1965). Preference, utility, and subjective probability. In D. Luce, R. Bus, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 249–410). New York: Wiley.

Luce, R. (1958). A probabilistic theory of utility. *Econometrica: Journal of the Econometric Society*, *26*(2), 193–224.

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, *24*(2), 178–191.

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception and Psychophysics*, *13*(2), 180–184.

Pagin, P. (2010). Vagueness and central gaps. In R. Dietz & S. Moruzzi (Eds.), *Cuts and clouds: Vagueness, its nature and its logic* (pp. 254–272). Oxford: Oxford University Press.

Pagin, P. (2011). Vagueness and domain restriction. In *Vagueness and language use* (pp. 283–307). London: Palgrave Macmillan.

R Development Core Team. (2016). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing Vienna Austria*.

Rayo, A. (2010). A metasemantic account of vagueness. In R. Dietz & S. Moruzzi (Eds.), *Cuts and clouds: Vagueness, its nature and its logic* (pp. 23–45). Oxford: Oxford University Press.

Rubinstein, A. (1988). Similarity and decision under Risk. *Journal of Economic Theory*, *46*, 145–153.

Schmidt, L. A., Goodman, N. D., Barner, D., & Tenenbaum, J. B. (2008). How tall is tall? compositionality, statistics, and gradable adjectives. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, *1*(5), 3151–3156.

Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, *23*(2), 113–128.

Solt, S. (2016). On measurement and quantification: The case of most and more than half. *Language*, *92*, 65–100.

Solt, S., & Gotzner, N. (2012). Experimenting with degree. *Proceedings of salt* (pp. 166–187).

Tafel, H., & Wilkens, A. (1963). Classification and quantitative judgment. *British Journal of Psychology*, *54*, 101–114.

van Rooij, R. (2011a). Implicit versus explicit comparatives. In P. Egre & N. Klinedinst (Eds.), *Vagueness and language use* (pp. 22–51). Basingstoke: Palgrave MacMillan.

van Rooij, R. (2011b). Vagueness and linguistics. In G. Ronzitti (Ed.), *Vagueness: A guide* (pp. 122–170). Dordrecht: Springer.

Weber, E. U. (1834). *De Pulsu, Resorptione, Auditu et Tactu. Annotationes anatomicae et physiologicae.* Leipzig: Koehler.

Wiener, N. (1914). A contribution to the theory of relative position. In *Proceedings of the cambridge philosophical society* (Vol. 17, pp. 441–449).

Wittgenstein, L. (1953). *Philosophische Untersuchungen/Philosophical Inves- tigations (translated by G.E.M Anscombe ed.)* (Anscombe ed.). Oxford: Blackwell.