UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
The Publicity of Concepts

Permalink
https://escholarship.org/uc/item/6pj5s7rd

Author
Misenheimer, Luke William

Publication Date
2014

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The Publicity of Concepts

by

Luke William Misenheimer

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Campbell, Co-chair
Assistant Professor Geoffrey Lee, Co-chair
Associate Professor Tania Lombrozo

Fall 2014

The Publicity of Concepts

Copyright 2014
by
Luke William Misenheimer

Abstract

The Publicity of Concepts

by

Luke William Misenheimer

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor John Campbell, Co-chair

Assistant Professor Geoffrey Lee, Co-chair

In academic philosophy and in ordinary life, it is often difficult to tell whether two people are really communicating, agreeing, or disagreeing about a single subject—about the metaphysical status of concepts, say, or about the color of a certain garment—or whether the two people are instead merely 'talking past' one another, respectively making claims about two different subjects. In attempting to make sense of these cases, it will often be helpful to ask the question of whether the two people are sharing a thought or are instead thinking different thoughts from one another. In many cases of interest, the distinction between true and merely apparent communication, agreement, or disagreement will come down to the distinction between sharing and not sharing a thought.

But although this analysis is fairly commonsensical, it presupposes that sharing thoughts is common. If sharing thoughts were somehow to turn out to be extremely rare, it would become difficult to make sense of the distinction between true and merely apparent communication *et cetera*, since if they involve sharing thoughts then these phenomena would themselves be extremely rare. The meaningfulness of this distinction and of many other distinctions we make in philosophy and in ordinary life requires that we often share thoughts with one another.

If concepts are the building blocks of thought, then sharing thoughts requires sharing concepts. So any theory of concepts that is intended to respect our ordinary way of thinking about ourselves should include an account of the publicity of concepts—an account of how it is that so many concepts are shared among so many different people. This enables the theory to give a reasonable analysis of everyday phenomena like communication, agreement, and disagreement.

However, there is a large class of theories of concepts that includes many popular contemporary theories and that is incompatible with the publicity of concepts. This is the class of 'individualistic' theories of concepts, so-called because they treat concept possession as something that only individuals do. These theories are incompatible with the publicity of concepts because they take many subtle features that distinguish individuals from one

another to be relevant to questions about which concepts those individuals possess, and so they inevitably answer "no" to any questions about whether a certain individual shares a certain high-level concept with another individual.

On individualistic theories, including the theories offered by Jerry Fodor and Christopher Peacocke, as well as others, the best picture that can emerge is one on which many individuals possess many concepts that are merely very similar to concepts possessed by others. But mere similarity cannot do the theoretical work required of concept sharing, so these individualistic theories are deeply incompatible with our ordinary way of thinking about ourselves.

The solution is to adopt a 'social' theory of concepts, which in contrast to an individualistic theory treats concepts as importantly socially articulated entities—as entities that have some social features at a more basic level than they have some other important features, such as their semantic values. Although social theories do face certain difficulties, it is possible to overcome these difficulties, in part by tying concepts very closely to public-language words and so accounting for the publicity of concepts together with the publicity of public languages like English. On balance, social theories should thus be preferred to individualistic theories if it is important to respect our ordinary picture of ourselves, because they allow accounting for the publicity of concepts but do not come with any serious, unavoidable theoretical costs.

Contents

Contents **i**

1 The publicity of thoughts **1**
1.1 Why care about publicity of thoughts 1
1.2 What sharing thoughts amounts to 8
1.3 How publicity of thoughts involves more than people sharing thoughts 15
1.4 What comes next 24

2 The problem of publicity **26**
2.1 Why individualistic theories cannot account for publicity 28
2.2 There is arbitrariness in the denotation of CAT, even if it is vague 29
2.3 Where there is arbitrariness, publicity needs explanation 42
2.4 Individualistic theories of concepts cannot explain how people share CAT . . 45
2.5 CAT is not special in being like this 48
2.6 Therefore, individualistic theories of concepts have a problem allowing for publicity 49

3 Causal and Inferential-Role Theories **52**
3.1 Causal theories of concepts 54
3.2 Inferential-role theories of concepts 68
3.3 Conclusion 81

4 Sameness and Similarity of Concepts **82**
4.1 Widespread sharing as a desideratum 84
4.2 Problems with understanding sharing in terms of similarity 90
4.3 Objections to understanding sharing in terms of sameness 104
4.4 Conclusion 116

5 A Social Theory of Concepts **117**
5.1 The state of play 118
5.2 Social theories of concepts 121
5.3 The theory 128

5.4 Consequences, applications, solutions 137
5.5 Public and private concepts 150
5.6 Conclusion 162

Bibliography **164**

Acknowledgments

I could not have written this dissertation without a great deal of assistance from many people, both inside and outside the UC Berkeley Philosophy Department. Special thanks are due to John Campbell and Geoff Lee for many long discussions and helpful comments on the content and form of this work in its many iterations; to Tania Lombrozo for comments on this work as well as for advising me on the qualifying exam topic that eventually grew into this project; to my fellow students Richard Lawrence and Jeremy Carey for joining me for many of the working sessions at which I actually wrote the bulk of this work; and to my spouse Cheryl Jerozal for patiently allowing me to probe her intuitions about many obscure topics.

Chapter 1

The publicity of thoughts

Although the ultimate goal of this project is to motivate a certain family of theories of concepts, I began this project out of an interest in distinguishing cases of true communication, agreement, and disagreement from cases in which two or more parties are merely 'talking past' one another—which in turn led to an interest in accounting for the publicity of thoughts. So in this introductory chapter, I will focus on that goal, explaining why it is important to account for the publicity of thoughts and what that publicity amounts to. Along the way, I will introduce some of the topics I will focus on in subsequent chapters, and I will try to give some sense of how those chapters fit into the project as a whole.

1.1 Why care about publicity of thoughts

I am interested in the publicity of thoughts because it seems to me that if important phenomena like communication, agreement, and disagreement are to be as widespread as we ordinarily take them to be, then thoughts must be public. Or, at least, if these phenomena as they are most naturally understood are to be as widespread as we ordinarily take them to be, then thoughts must be public. So I want to account for the publicity of thoughts because of the confluence of two reasons, which I will briefly present in this section and then discuss at greater length in later chapters. The first reason, which I will take up again in chapter 4, is that it would be nice if thoughts were public, since that would allow us to hold on to ordinary understandings of phenomena like communication, agreement, and disagreement. The second reason is that if thoughts are indeed public, then this fact needs to be specially accounted for, since it is at odds with some common ways of thinking about thoughts—we cannot take the publicity of thoughts for granted. This second reason will be the focus of an argument that I will develop throughout chapters 2, 3, and 4.

1.1.1 Communication, agreement, and disagreement require sharing thoughts

In saying that communication, on an ordinary understanding, requires publicity of thoughts, I do not mean to imply that there is a canonical, obvious, fully-detailed, easily-articulated pre-theoretical understanding of communication. There is, of course, no such thing. If there were, the history of philosophical analysis of the phenomenon would be very different than it actually is. However, there is still some ordinary understanding of communication that (among other things) informs our analyses of the phenomenon, and any theory that is inconsistent with this understanding comes with a philosophical cost: all else being equal, we should prefer theories that allow us to hold on to (or revise as little as possible) this ordinary understanding.

And even if the ordinary understanding of communication does not always give determinate answers to questions about what counts as communication, it at least requires that in successful cases of communication, the communicator and the audience must share a thought. There must be something—the *message*, that which is communicated—that the communicator wishes to get across to the audience and that the audience comes to think (or perhaps believe or understand) as a result of the communicative act. If, as a result of the communicative act, the audience only comes to think something that is not quite what the communicator had in mind, then the communication was not entirely successful. For example, if A says to B "There's a cat ahead," prompted by the thought that there is a cat ahead, but B only comes to believe that there is a *furry* cat ahead, then at best A only partially succeeds in communicating.

I hope it is fairly intuitive that, in the case just described, A did not totally succeed in communicating. Just in case it is not intuitive, here is one reason for understanding the case in the way I have suggested: Ordinarily, we take successful communication to be a way of transmitting knowledge from the communicator to the audience. Supposing that A is attempting to communicate something A knows, then, B should come to have knowledge if A succeeds in communicating. But what A knows is just that there is a cat ahead, which would be true even if there were a furless cat ahead. And since B comes to believe that there is a furry cat ahead thanks to A's communication, the truth of B's belief is not guaranteed by what A knows—given only that A knows that there is a cat ahead, it is left to luck whether the cat is furry (and so B is right) or the cat is furless (and so B is wrong). Since B's getting it right would be a matter of luck, B does not seem to gain knowledge by coming to believe that there is a furry cat ahead on the basis of A's communication. Since no other special knowledge-preventing circumstances held in the case, A's communication must not have been totally successful. If communication is to be a medium for transmission of knowledge, it must also be a medium for transmission of thoughts—and so it must involve sharing thoughts.[1]

[1]There are, of course, cases in which B comes to know something on the basis of A's testimony when the two do not share thoughts. For example, if A says "There's a furry cat ahead" and B, not being familiar with the word "furry" (but knowing that it is not some kind of negation), only comes to believe that there

This feature of our ordinary understanding of communication—that communication requires sharing a thought—can also be seen when it shows up in philosophical accounts of the phenomenon. For example, according to many accounts, communication involves the communicator having a special intention that the audience come to have a certain thought (or, more specifically, belief). On these intention-based accounts, the communicator communicates that p just in case she has a certain intention that, among other things, the audience come to think that p. If the success of communication can be identified with the fulfillment of this intention, then on these accounts successful communication that p requires that both parties have the thought that p. Other accounts of communication have it involving a special conventional association between a certain kind of act and a certain thought: the communicator communicates that p just in case her action is associated in a certain way, by convention, with the thought that p. On these convention-based accounts, communication might not require that a particular communicator share a thought with her audience, but it is hard to see how convention could associate an act with the thought that p without many people sharing the thought that p, or at least being able to share the thought that p.[2]

So on an ordinary understanding of communication, which it would be nice not to have to reject, successful communication requires sharing thoughts between individuals. And since, on an ordinary understanding, cases of successful communication are fairly common among humans in the actual world, an ordinary understanding requires widespread sharing of thoughts. So a theory of thoughts that allows for widespread thought sharing is, all else being equal, preferable to a theory that does not allow widespread thought sharing among humans in the actual world—at least if it is important that the theory respect our ordinary understanding[3].

And what goes for communication goes for agreement and disagreement as well. On ordinary understandings of the phenomena, two people agree with one another just in case they both believe or judge the same thing, and they disagree with one another just in case one dissents from something the other assents to.[4] For both phenomena, the two people

---

is a cat ahead, it is possible that B's belief counts as knowledge, even if the thought that there is a cat ahead never occured to A (because, say, A is preoccupied with furriness). In general, if B forms a belief with weaker truth conditions than A's belief and assertion, then if A's belief counts as knowledge B's will also—though there are many counterexamples to this general principle. But this cannot be the standard for cases of knowledge transmission by communication, since we cannot count on the audience's belief always being weaker than the communicator's belief. So the standard must involve sharing thoughts.

[2]Paul Grice (1957, 1968, 1969) offers perhaps the most influential intention-based account on which communication requires sharing thoughts, and David Lewis (1975) perhaps the most influential convention-based account of communication, or at least of linguistic communication. Accounts following in these traditions include that proposed by Brian Loar (1981) and that proposed by Stephen Schiffer (1972) but later rejected by him (1987).

[3]For now, I will simply assume that it is important for at least some theories to respect our ordinary understanding. I will discuss this subject in detail in chapter 4.

[4]Disagreement might carry a bit less argumentative weight here than the other two phenomena of interest (communication and agreement). This is because in the case of disagreement, there is a somewhat viable alternative account which does not require shared thought: it might be that two people disagree with one another just in case one has a belief, the other has a belief, and it cannot be that both beliefs are true. John

involved must share a thought—in one case they must both believe or judge it, and in the other case it must be that one believes or judges it while the other disbelieves or otherwise dissents from it. And again, both phenomena are ordinarily understood to be fairly common, so in order to allow us to hold on to our ordinary understandings of these phenomena, a theory of thoughts must allow widespread thought sharing.

This, then, is an important sense in which thoughts might be 'public': it might be a fairly common occurrence for pairs of humans to share thoughts. In this work, I will be primarily concerned with this kind of publicity and with a stronger kind that I will introduce in section 1.3. I will present some other kinds of publicity in which I am less interested in section 1.2, and I will also say more then about what I mean when I talk about sharing thoughts.

So any theory that does not allow thoughts to be public in at least the sense that they are actually shared fairly often incurs a serious cost. Of course, it might be that in order to allow for this kind of publicity, a theory would have to incur other costs which would outweigh the benefits of allowing us to hold on to ordinary understandings of communication, agreement, and disagreement. Over the course of the remaining chapters of this work, but especially in chapter 5, I will investigate what it takes for a theory of thoughts to allow widespread thought sharing among humans in the actual world, as well as what it takes to allow for thoughts being 'public' in a thicker sense. In chapter 4, I will also discuss at greater length the costs associated with denying publicity. In particular, I will discuss the prospects for giving alternative accounts of phenomena like communication, agreement, and disagreement—accounts on which these phenomena do not involve sharing thoughts.

1.1.2 Other important phenomena also require publicity of thoughts

There is, of course, much more to our lives as thinking, social creatures than just communication, agreement, and disagreement. My focusing on these phenomena is more a matter of personal preference than a principled philosophical stance—I became interested in the publicity of thoughts because of my interest in philosophical debates in which it was unclear to me whether the parties involved were actually successfully communicating, agreeing, and disagreeing with one another or were instead merely 'talking past' one another. I certainly would not wish to suggest that communication, agreement, and disagreement are the only

MacFarlane (2007) offers an account in this style as a potential improvement over the "obvious" account of one party accepting something the other party rejects, but his motivations for this potential improvement fall outside the scope of this project. In any case, I still think disagreement is worth mentioning, since this alternative account might not be entirely satisfactory, and even if it is satisfactory it will not, in the end, obviate the need for the publicity of thought. This is because even on this alternative account, in most of the interesting real-world cases disagreement will still require a very close relation between the truth conditions of the two beliefs in question, and it will be unreasonable to expect this close relation on a picture on which thought is not public. Why this is will become more clear over the next three chapters, so that in chapter 4 when I return to this point, it should be fairly obvious.

important phenomena which are most easily understood as involving thoughts which are public in some sense, as a few other such phenomena come readily to mind.

Though it is surely related to communication, as discussed above, the transmission of knowledge from one individual to another is worth mentioning on its own as a phenomenon requiring the sharing of not just thoughts but beliefs more specifically. In normal cases, B can come to know that *p*, say, thanks to the testimony of A only when A herself knows that *p*. If A only knows some very similar fact, the fact that *p'*, then if B cannot infer *p* from *p'* there is little hope of B coming to know that *p* thanks to the testimony of A. There are, of course, counterexamples to this general principle, including the case of B coming to know *that A testified that p'*, but the existence of these cases does not falsify the claim that really matters here, which is that the transmission of knowledge by any means is an important phenomenon which involves the sharing of thoughts.[5]

It is also worth mentioning our ability to sympathize and empathize with others and our ability to predict and explain others' actions based on our own knowledge of their beliefs and desires. For example, if B is to understand how A feels upon learning that there is a cat ahead, or if B is to base her predictions about what A will do on the fact that A believes that there is a cat ahead—or at least if B is to do either of these things in the way that normal humans do—then B must herself be able to have the thought that there is a cat ahead.

1.1.3 Publicity of thoughts cannot be taken for granted

I hope it is by now somewhat clear why I think that a theory of thoughts that allows thoughts to be public is better, other things being equal, than a theory that does not allow thoughts to be public—at least in the minimal sense of 'public' that requires thoughts to be shared often. But even if this is true, why should it substantially affect our theorizing? Why can't we simply add to our theories an assumption that thoughts are public—or why should we do any more than refrain from adding to our theories any statements which are obviously inconsistent with thoughts being public?

The problem is that many popular ways of thinking about thoughts are already inconsistent with the publicity of thoughts, even if this inconsistency is not immediately obvious. I will spend all of chapters 2 and 3 discussing the barriers to publicity thrown up by existing theories and the prospects for minimally amending those theories to allow for the publicity of thoughts. For now, though, I want to give some sense of the problem I have in mind by discussing one particular theory which is more obviously inconsistent with the publicity of

[5]This line of reasoning deserves more discussion than I can give it here, since there are many details I have ignored. For example, there might be further doubts about whether coming to know something by testimony generally requires that the testifier knows what the audience comes to know, since there might seem to be some more important cases in which, thanks to testimony, the audience gains knowledge which the testifier lacks. When discussing this issue, Jennifer Lackey (1999) offers some interesting cases that seem to have this feature. However, I think many of the grounds one might have for doubting that communication is a medium for knowledge transmission—in particular, many of the grounds offered by Lackey—will not turn out to be grounds for doubting that communication is a medium for thought transmission, so much of the discussion of these issues will not be directly relevant to this project.

thoughts than some others. Neither the theory I will present nor the complaint I will make against it is new, but both taken together are illustrative of the problem I am trying to overcome in accounting for the publicity of thoughts.

The theory that I wish to present is really a theory (or part of a theory) of concepts, not a theory of thoughts. I will say a bit more about how I wish to understand thoughts, concepts, and the relations between them in section 1.2. For now, though, I will simply assume that sharing thoughts requires sharing the concepts involved in those thoughts, and so any serious obstacle to sharing concepts will be a serious obstacle to the publicity of thoughts. The theory I will present is inconsistent with the publicity of thoughts, given this assumption, because it implies that almost no one shares any concepts with anyone else.

So consider a theory that involves the claim that in order to share a concept, two people must be inclined to make all the same inferences as one another involving thoughts involving that concept. This might seem like a fairly reasonable answer to the question of what it takes to share a concept: after all, a concept's role in inference might seem as good a candidate as any to be an individuating feature, and for many concepts it seems completely arbitrary to select just some of the inferences they feature in as the really important ones, so in order to avoid arbitrariness one might take all of the inferences a concept features in to distinguish it from all other concepts.[6]

To take an example, then: If A is inclined to use the concept CAT in the inference from the thought that Felix is ahead to the thought that there is a cat ahead[7], but B is not, because B does not know Felix, then on this theory B does not share the concept CAT with A, and so (because sharing thoughts requires sharing the concepts involved in those thoughts) B cannot share the thought that there is a cat ahead with A. B must instead express some different thought in saying "there is a cat ahead." We might call B's thought "the thought that there is a cat' ahead," instead of "the thought that there is a cat ahead." However, on this theory we might end up wanting to differentiate the two thoughts even more, since there will be barriers to A and B sharing many other concepts as well, such as the concept AHEAD.

Indeed, it is hard to come up with any concepts at all that A and B would share on a theory like this. For instance, even if A and B know all the same dogs, and seem to be acquainted with all the same facts about dogs, they would not share the concept DOG if, say, A is inclined to use the concept DOG in the inference from the thought that Fido is a dog to the thought that Fido is not a cat. This is because B cannot possibly make the same inference, because B cannot have the thought that Fido is not a cat (or at least the same thought that A has that we might call "the thought that Fido is not a cat"), because B does not have the concept CAT, because B does not know Felix.

[6]Though their motivations are certainly more worthy of consideration—and the theories they end up with are more sophisticated—than the cartoon I have presented here, philosophers who have proposed theories like this include Ned Block (1986) and Robert Brandom (1994).

[7]That A is inclined to 'use' the concept CAT in this inference just means that A is inclined to make this inference and at least one of the thoughts involved in this inference itself involves the concept CAT.

So, on a theory which individuates concepts according to their total role in inference, it is hard to see how any two people could ever share concepts.[8] This observation might just motivate a theorist to individuate concepts according to some other features of the ways they are used, such as their causal relations to properties in the world, or some very small part of their role in inference. However, as I will argue in chapters 2 and 3, such a theorist would still be rejecting publicity, just less obviously than one who individuates concepts according to their total role in inference. This kind of rejection—rejection of publicity by acceptance of some otherwise reasonable claims about the mind that are ultimately inconsistent with publicity—will be the main obstacle to publicity that I will try to overcome in the chapters that follow.

1.1.4 Sharing thoughts involves more than having similar thoughts

In addition to giving a preview of the main obstacle I will be trying to overcome with the picture I develop, it is perhaps worth giving a preview of the picture of humans as thinking social creatures that I am trying to avoid—the picture I am most worried will develop if the obstacle is not overcome.

Suppose that Crusoe and Friday are alone on an island together, but the two have such different histories and dispositions that they conceptualize the world in very different ways and cannot truly communicate with each other. Each might still guess that the other wants something like shelter (say, if Crusoe possesses the concept $\text{SHELTER}_{\text{C}}$ and Friday possesses the concept $\text{SHELTER}_{\text{F}}$, and both of these idiosyncratic concepts are in some sense similar to SHELTER but have slightly different denotations), and the two might still work together to erect a shelter (an object that turns out also to be a $\text{shelter}_{\text{C}}$ and a $\text{shelter}_{\text{F}}$). But even if both makes guesses about the other's mind that allow the two to loosely coordinate their actions, their guesses will never be both precise and correct.[9] The two might try to really agree or disagree about some point of business, to communicate, or to figure out what the other is thinking, but they will never really succeed, because they do not share the concepts they would need in order to share thoughts. The closest they could come to sharing thoughts would be having thoughts that are in some sense very similar but still not the same.

The picture of humans as thinking social creatures that I am most concerned to avoid is one on which we all actually turn out to be like Crusoe and Friday in the example above. I am most concerned to avoid this picture in part because I can see its attraction. It is not without motivation: Because we humans are really quite similar to one another in many

[8] Jerry Fodor (2004) makes this complaint very forcefully, and my aim here is simply to echo him. However, in chapter 3 I will make the an analogous complaint against Fodor's own theory of concept individuation, as well as several others.

[9] The most straightforward way, in the case described, either party could make a correct guess about the other's thoughts would be to make a deliberately imprecise guess. For example, if Friday guesses that Crusoe means to erect *something like* a $\text{shelter}_{\text{F}}$, then Friday would be correct if (as only someone possessing both concepts $\text{SHELTER}_{\text{C}}$ and $\text{SHELTER}_{\text{F}}$ could explicitly know) a $\text{shelter}_{\text{C}}$ is indeed something like $\text{shelter}_{\text{F}}$.

ways, and because we all live in the same wider world, it would make sense to expect that most of us cannot help but think about the world in many ways which are very similar to ways others think about it. And such similarity would be enough to explain many social phenomena, including those required for loosely coordinated action. Why hope for anything more?

So far, I have only given one reason to prefer a picture on which we are more closely connected to one another than we are on this picture I wish to avoid. That reason is that this picture would force us to radically revise our ordinary understandings of communication, agreement, and disagreement. It would do so because our ordinary understandings of those phenomena involve different people having the same thought, and this picture I wish to avoid only has room for different people having very similar thoughts. Having very similar thoughts is not a satisfactory theoretical substitute for having the same thought for several reasons, which I will discuss when I take up the prospects of this picture at greater length in chapter 4. For now, I will just briefly mention two: firstly, having very similar thoughts cannot do much of the theoretical work having the same thought is called on to do, since "very similar" supports so few of the important inferences that "the same" does (Perhaps most obviously and importantly, similarity is not a transitive relation, while sameness is.); secondly and relatedly, I am doubtful that much sense can be made of the similarity of thoughts in cases in which there is no hope of true sameness of thoughts[10].

1.2 What sharing thoughts amounts to

In the last section, I introduced the first of two kinds of publicity of thoughts for which I am interested in accounting. I will discuss the other, more robust, kind in the next section, but first I would like to spend this section giving a clearer idea of the first kind, which I will call "minimal publicity":

minimal publicity: Minimal publicity holds just in case it is common for pairs of individuals to share thoughts.

I will contrast minimal publicity with the more robust kind of publicity I am interested in, which I will call "robust publicity," as well as other phenomena which might be called "publicity of thoughts" but which involve more or less than minimal publicity does.

I am interested in minimal publicity because I think it is clearly required if we are to hold on to anything like our ordinary understanding of ourselves as thinking social creatures, as I discussed briefly in the last section and will discuss at greater length in chapter 4. As will come out in the next section, I am interested in robust publicity because I think it is also required—though perhaps not as clearly—if we are to hold on to our ordinary understanding of ourselves as thinking social creatures. In addition to that, though, I am interested in robust

[10]Jerry Fodor argues for a point like this about similarity and sameness of concepts in *Concepts* (1998), and I see no reason why his argument could not be adapted to work for thoughts.

publicity because I think there is little hope for accounting for minimal publicity without accounting for something like robust publicity along the way. This is because, as I will argue in the remaining chapters of this work, the best hope of accounting for minimal publicity comes with adopting what I will call a 'social' theory of concepts; and such a theory brings along with it an account of robust publicity.

My goal in this project will be to account for whichever of these two kinds of publicity I can. I hope that even if my attempt to account for robust publicity is ultimately unmotivated or unsuccessful, my attempt to account for minimal publicity will still be interesting in its own right.

So then, on to minimal publicity. I will not be offering any serious analysis of "common" or "individuals" as they show up in the formulation of minimal publicity above, since I doubt that many serious issues will turn on how they are understood. However, "share thoughts" requires some discussion, both for the sake of clarity about the goals of the project at this early stage and in order to lay the foundations for the arguments to come.

1.2.1 Sharing thoughts requires having thoughts with the same truth conditions

My interest in the publicity of thoughts is driven by my desire for philosophical theories of the mind that remain in touch with folk psychology, so it should not be surprising that the 'thoughts' mentioned in the publicity of which I am interested are entities described by theories grounded in folk psychology. I am interested in thoughts just as entities importantly involved in phenomena like communication, agreement, and disagreement—and so also knowledge, belief, assent, dissent, desire, judgment, understanding, consideration, and many others. I am interested in a thought as what is communicated or what is agreed to, and what is communicated or agreed to can also be known, believed, denied, desired, judged, understood, considered, et cetera. What is communicated, et cetera, is also often importantly either true or false, so I am also interested in thoughts as truth-apt entities.

There are, of course, many different philosophical theories that have roughly this starting point.[11] I hope that this project will be of interest to adherents of many such theories, since many of the details of the theories will not be relevant to this project. For example, some theorists might disagree about whether thoughts are best understood as abstract entities or as concrete entities, or about how much structure thoughts have necessarily. I think I can safely remain agnostic with respect to many of these debates.

But some points concerning thoughts are very important to my project. In particular, my project depends a great deal on what exactly is involved in sharing thoughts, since that is involved in turn in the publicity of thoughts. So: what is it to share a thought, and what

[11] There are also, of course, many theories that do not assign any real importance to the truth conditions of thoughts. However, the debate over what if any role should be assigned to the truth conditions of thoughts is well outside the scope of this project. I am simply assuming that thoughts are importantly truth-apt entities here.

does it require? One folk-psychological gloss on "sharing a thought" is "thinking the same thing," but this notion must be made a bit more explicit if it is to do any work in more rigorous theories. And how it should be made more explicit depends on which kind of theory it is meant to do work in.

Depending on whether thoughts are abstract entities or concrete entities, A and B sharing a thought amounts either to A having a thought which is one and the same thought as a thought B has or A having a (token) thought which is of the same type as a (token) thought B has. The important theoretical question which underlies the question of whether A and B share a thought, then, is the question of how thoughts are individuated (if they are abstract entities) or typed (if they are concrete entities). From this point on, for the sake of readability, I will mostly talk about thoughts as abstract entities, taking up the question of how they are individuated rather than the question of how they are typed. All that I have to say about thoughts, though, should be easily translatable into the language of a theory which treats thoughts as concrete entities.[12]

So: how are thoughts individuated? One plausible answer is that they are individuated at least in part according to their truth conditions. For example, if A has the thought that there is a cat ahead, and B has the thought that there is a furry cat ahead, then one reason I might cite in explaining why A and B are thereby thinking different things is that there is a possible situation in which A's thought would be true and B's thought would be false, namely the situation of a furless cat being ahead. Since one of my starting points was an interest in thoughts as truth-apt entities, it is reasonable to say that any two thoughts with different truth conditions are really two distinct thoughts and not really one and the same thought.

This is hardly a complete answer to the question of how to individuate thoughts, though. For instance, it does not settle the question of whether thoughts have some kind of form, sense, intension, or mode of presentation which also individuates them, separately from their truth conditions—whether, say, the thought that Hesperus is visible is distinct from the thought that Phosphorous is visible, or whether the thought that there is a black cat ahead, thought by someone who has seen blackness for themselves, is distinct from the thought that there is a black cat ahead, thought by a congenitally blind person who only knows of blackness from the testimony of others. It also does not settle the question of just what kinds of truth conditions are relevant to individuating thoughts—whether, say, the truth conditions could be modeled just using possible worlds, so that A's thought at noon that there is a cat ahead (of me now) is different from B's thought at midnight that there is a cat ahead (of me now), or whether the truth conditions could only be modeled by more complex entities, such as world-center-time triples.

However, even though it is a very incomplete answer to say that thoughts are individuated at least according to their truth conditions, it is the only answer I wish to give. The barriers

[12]For example, when translated into the language of a concrete-thoughts theory, the question of how to individuate thoughts becomes the question of how to type thoughts or the question of how to individuate thought types, and the question of what it is to have a certain thought becomes the question of what it is to have a token thought of a certain type.

to sharing thoughts that would be thrown up by also individuating thoughts according to their modes of presentation or by modeling truth conditions just using possible worlds are interesting and important. In the next subsection, I will discuss some fairly minimal respects in which I will take thoughts to be individuated according to more than just their truth conditions. Unfortunately, though, the scope of this project must be limited if it is to have any chance of success. So in this work, I will focus on the barriers to sharing thoughts that are thrown up by individuating thoughts according to their truth conditions, and I will not concern myself with any barriers that are thrown up by the indexicality of thoughts. That is, I will try to account for the publicity of thought from within the constraints imposed by making the following assumption:

individuation by truth conditions: If A has the thought *t*, and B has the thought *t'*, then if there is a possible situation (which we might model as a world with a center at a time, or perhaps some more complex entity) in which *t* and *t'* would have different truth values, A and B are not sharing thoughts by respectively having *t* and *t'*. (Alternatively: If A and B are sharing thoughts, then there is some thought which A is having, *t*, and some thought which B is having, *t'*, such that there is no possible situation in which *t* and *t'* would have different truth values.)

As I will show in chapters 2 and 3, these constraints are substantial.

1.2.2 Sharing thoughts also requires sharing concepts

Though the constraints imposed by individuating thoughts at least according to their truth conditions are substantial, they are somewhat unwieldy. So, for the sake of argumentative simplicity, as well as for the sake of bringing the project into closer contact with the various theories I will be discussing, I would also like to introduce another assumption:

individuation by concepts involved: If A has the thought *t*, and B has the thought *t'*, then if there is some concept which is involved in *t* but not *t'*, then A and B are not sharing thoughts by respectively having *t* and *t'*. (Alternatively: If A and B are sharing thoughts, then there is some thought which A is having, *t*, and some thought which B is having, *t'*, such that there is no concept which is involved in *t* but not *t'*.)

Given that one cannot have a thought that involves a concept if one does not possess the concept, this assumption brings with it a useful corollary:

dependence of sharing thoughts on sharing concepts: If A and B do not share a certain concept, they cannot share any thoughts involving that concept.

Both of these claims, though, require significant explanation. In particular, if I want either of these claims to do any important work, I need to make clear how they are to be understood by explaining how I wish concepts to be understood for the purposes of this discussion.

As was the case with thoughts, I would like my talk of concepts to import to the discussion as little objectionable theoretical baggage as I can get away with. Although I will be using a fairly substantive understanding of concepts, I hope that relatively few readers will find my picture incomprehensible, and I hope that it will be basic enough that those with incompatible understandings will still be able to translate what I have to say about concepts into language they find less objectionable.

However, I will be taking concepts as in many ways explanatorily prior to thoughts. The explanatory importance I attach to concepts is part of the foundation of this project, so unfortunately much of what I have to say will not be relevant to theories on which there is no sense in which thoughts have conceptual structure—for example, theories on which the only interesting structure thoughts have is that of sets of centered possible worlds, or thoughts are merely relations to public-language sentences—or to theories on which thoughts are entirely explanatorily prior to concepts—for example, theories on which concepts are merely useful theoretical abstractions that have no real existence.

So: the most basic theoretical role I take concepts to play is the role of constituents of thoughts that help to explain some of the important features of the thoughts they make up. Most importantly, concepts are entities involved in thoughts that help determine the truth conditions of the thoughts they are involved in by (perhaps among other things) denoting or referring to certain objects or properties, and concepts are entities the possession of which enables thinkers to have thoughts involving them.

For example, one difference between the thought that there is a cat ahead and the thought that there is a dog ahead is that the former is a thought about cathood or cats and the latter is not. One explanation of this difference is that the former thought involves the concept CAT, denoting cathood or cats, and the latter does not—someone must use the concept CAT in thinking the former thought, but they need not do so in thinking the latter. And only someone who possesses the concept CAT is able to use that particular concept in thought.[13] Concepts, then, are most important in the way they relate thinkers who possess them to particular objects and properties: CAT enables a thinker who possesses it to think about cathood or cats, DOG doghood or dogs, WATER water, RED the color red, HESPERUS the planet Venus,[14] and so on.

[13]As with thoughts, depending on one's theory of concepts, it might be more convenient to take concepts to be abstract entities, making the relevant question the question of how concepts are individuated, or it might be more convenient to take concepts to be concrete entities, making the relevant question the question of how concepts are typed. As with thoughts, when discussing concepts I will be using the language of individuation, but it should be straightforward enough to translate what I have to say into the language of typing.

[14]It might seem strange to talk about 'concepts' of particular individuals. In many theories, "concept" is reserved for entities which only denote properties or kinds and never refer to individuals—in many theories, concepts are necessarily general in their application. I wish to use "concept" somewhat more broadly here, as a term for any entity the possession of which enables someone to think about something, whether that 'something' is general or particular. However, my discussion will largely focus on general concepts, so I hope it will still be intelligible to theorists who reserve "concept" for entities which denote properties or kinds.

On this understanding, there is real explanatory work being done by concept use and concept possession, and in a full theory these phenomena would need their own explanations in turn. However, I do not wish to present a full theory of concepts here. Concerning concept use, all I would like to say here is that using a certain concept is involved in having a thought which involves that concept. To say more would be to take a side in debates that are not directly relevant to my project.

I also prefer not to say a lot about concept possession now, in part because I will have much more to say about it in the remaining chapters. But I do need to say something if I want to head off an objection to the way I introduced concepts to the discussion. I introduced them by suggesting that once the assumption has been made that thoughts are individuated at least in part by their truth conditions, it should be easy to accept the assumption that thoughts are individuated at least in part by the concepts they involve and the corollary that sharing thoughts depends on sharing concepts.

In part because concept possession is a fairly thick notion on my understanding of concepts, it is not obvious why anyone who accepts the assumption that thoughts are individuated at least according to their truth conditions should also accept the assumption that thoughts are individuated at least according to the concepts they involve. There is room to accept that thoughts are individuated by truth conditions but deny that they are individuated by concept involvement, simply because there is room to deny that there is at most one concept for each object or property that thoughts might be about. In other words, there is room to deny that concepts are individuated only according to their contributions to the truth conditions of thoughts in which they are involved—to deny that concepts are individuated only according to what we might call their 'semantic values.'

And it would be quite reasonable to deny that concepts are individuated only according to their semantic values, if concepts are meant to be entities with real theoretical weight. For example, it is hardly obvious that the concepts WATER and H2O have different semantic values (at least in the framework I am using here), but it seems quite reasonable to claim that they are different concepts, and the same could be said of the concepts HESPERUS and PHOSPHOROUS. We might describe the difference between these concepts as a difference in their senses, intensions, or modes of presentation, where these are understood as conceptual features that are distinct from semantic values. And if, say, we individuate concepts according to their modes of presentation as well as their semantic values, then we leave room for thoughts that have the same truth conditions but that involve different concepts—for example, the thoughts that there is water ahead and that there is H_2O ahead, or the thoughts that Hesperus is visible and that Phosphorous is visible. The dependence of sharing thoughts on sharing concepts, then, would not simply follow from individuating thoughts according to their truth conditions.

Individuating concepts according to more than just their semantic values clearly creates extra barriers to publicity beyond those created by individuating concepts according to their semantic values. As was the case with thoughts, though, because the scope of this project must be limited if it is to have any hope of success, in the rest of the discussion I will simply ignore any barriers to publicity beyond those created by individuating concepts

according to their semantic values, and so I will ignore the—surely very reasonable—options of individuating concepts according to their senses, intensions, or modes of presentation. This ignorance will allow me to make the useful assumption that sharing thoughts depends on sharing concepts, even though that assumption clearly does not follow from the other assumptions I have made.

Because of these limits on the scope of the project, the best I can hope for is to succeed in explaining how thoughts are public as far as their truth conditions are concerned. While this will leave some work to be done in fully accounting for the publicity of thoughts—in overcoming the barriers to publicity I will be ignoring for now—I think there is some hope that the framework I develop for accounting for the publicity of thoughts in the limited domain of truth conditions might be extended to cover the publicity of thoughts in other domains as well, such as senses or modes of presentation.

1.2.3 There are also other senses in which thoughts might be public

The last thing I will do in this section to try to give a clear understanding of minimal publicity is to contrast it with a couple other phenomena which might be called 'publicity of thought' and which might also be of philosophical interest.

First, an important initial step in accounting for the publicity of thoughts is giving an account of thoughts on which thoughts are the kinds of things which it is even possible to share with others—on which thoughts are not necessarily private. For example, if an individual's thoughts are understood to be particular episodes in that individual's mind which in principle cannot be compared in any meaningful way to the particular episodes in any other individual's mind, then there is no hope for any kind of publicity of thoughts. It is important, then, for anyone who cares about the publicity of thoughts to start with an understanding of thoughts on which one individual's thoughts can be meaningfully compared to another's.

There are various ways to allow for this kind of publicity, the possibility of meaningful comparison among different individuals' thoughts. One might understand thoughts as concrete entities which are typed according to their truth conditions, for instance, or as abstract entities which can somehow be 'grasped' by different individuals.[15]

I will not be focused on this kind of publicity in this project, but I do not consider it to be unimportant. Instead, I will not be focused on it because I would like to think that the problem of allowing for this kind of publicity—the problem of across-person in-principle comparability of thought—has already been solved. I will focus on the problem of allowing for minimal and robust publicity because I consider it to be an important problem which arises once the problem of in-principle comparability has been solved.

[15]For example, concern for this most basic kind of publicity seems to have motivated Gottlob Frege (1956) to distinguish thoughts as abstract entities from ideas as mental entities.

Second, it is perhaps worth mentioning a phenomenon which might be called 'publicity of thought,' or which at least involves some kind of publicity of thought: A thought might be public in the sense that everyone in a community of two or more individuals intends it (or has an intention involving it), believes it, or knows it. This kind of publicity—or even stronger kinds[16]—might be required for other important social phenomena, such as collective actions, conventions, and some cases of rational coordination. However, I will not be focusing on phenomena like this in this project because I take them to depend on minimal publicity, if not something stronger, and to involve more than it: in order for it to be common for everyone to believe a certain thing, it must be common for multiple people to be able to think the same thing, since believing involves or is a kind of thinking. I have enough work ahead of me just trying to account for the two kinds of publicity I am interested in without also accounting for these stronger phenomena.

1.3 How publicity of thoughts involves more than people sharing thoughts

In section 1.1 I tried to present the main reason I am interested in minimal publicity—in it being common for different people to share thoughts. Briefly, the reason is that on our ordinary understandings of phenomena we take to be parts of our lives as social, thinking creatures—phenomena like communication, agreement, and disagreement—these phenomena require that different people share thoughts. I hope this is a relatively uncontroversial claim, since it is the minimal foundation on which I can construct this project.

However, in this section I will try to expand that foundation by offering some reasons to be interested in something more than minimal publicity, something I will call 'robust publicity.' These reasons will themselves again be founded on our ordinary understandings of phenomena like communication, agreement, and disagreement, but I expect they will be somewhat more controversial than the reason I presented for being interested in minimal publicity. So even though robust publicity will turn out to be an important goal for this project, I began by trying to motivate an interest in minimal publicity in the hopes that I might keep the interest of readers who do not share my motivations for being interested in robust publicity, since accounting for minimal publicity will remain a secondary goal.

I will now proceed to give the reasons for my interest in something beyond minimal publicity, and then I will try to give some sense of what this more robust publicity might be like. This will not be a fully detailed picture of robust publicity, since that will only come into view in chapter 5, once I have completed more of the groundwork for the account of robust publicity. However, even at this first stage of the project I think it will be helpful to get some idea of where things are headed. Then, once I have given a brief description

[16]For example, a thought that p might be public knowledge in a community in the sense that each member of that community knows that p, knows that everyone knows that p, knows that everyone knows that everyone knows that p, and so on.

of robust publicity, I will compare it to some closely related phenomena discussed by other authors.

1.3.1 Communication, agreement, and disagreement require more than simply sharing thoughts

Quite simply, I am interested in robust publicity because I think there are several ways in which, on our ordinary understandings, our lives as social, thinking creatures involve more than the mere actuality of our sharing thoughts. In particular, we take ourselves to be able to share thoughts across a wide range of circumstances, allowing for a wide range of social interactions. Some theories of thoughts that allow thought sharing in certain circumstances do not allow thought sharing across this wide range of circumstances, and so some theories of thoughts that allow something like minimal publicity do not allow thoughts to be as public as we might want.

For instance, the theory of concepts mentioned above—on which concepts are individuated at least according to their total role in inference—can accommodate thought sharing in some circumstances, in particular in circumstances in which all the inferences one person is disposed to make, considered as an entire system, are perfectly isomorphic to all the inferences which another person is disposed to make. However, this is an extremely narrow range of circumstances when compared to the range across which we might like to allow thought sharing. So although this theory of concepts is not incompatible with any thought sharing, the restrictions it places on thought sharing are still much too strong, since it does not allow any thought sharing among any actual humans.

That sharing takes place among actual humans is an obvious requirement imposed by our ordinary understandings of communication, agreement, and disagreement. I have already explained that these phenomena, on our ordinary understandings, involve sharing thoughts. And it should go without saying that they also take place among actual humans. In the rest of this subsection, I will present several other requirements that seem to me to be imposed by our ordinary understandings of communication, agreement, and disagreement—requirements on what is acceptable in any theory that has implications for the sharing of thoughts. The requirements will mostly amount to requirements that the theory must allow sharing to have certain features.

First, transitivity of sharing: on an ordinary understanding of the phenomenon, communication requires a kind of sharing of thoughts that is transitive. Communication does not occur only between isolated pairs of people, but along chains of people. If A successfully communicates some message to B, then B should be able to pass that same message along to C, and C to D, and so on, such that even after a long chain of communication, if every instance is successful then Z can come to think what A thought. Of course, this is not to say that there are no long chains of communication in which a message changes by the end. But if each individual in the chain really does pass on whatever message she received, then if the message has changed by the end then the chain of communication must not have been

entirely successful.[17]

Second, sharing among many individuals: on ordinary understandings of the phenomena, communication, agreement, and disagreement require thoughts to be shared among large groups, not just between individuals. Someone giving a speech might succeed in communicating their thoughts to thousands or millions of others. Every member of a large group might agree with every other member of the group and disagree with every member of another group on some tenet central to group membership. These phenomena involve a single thought being shared by many individuals, and they do not require each person sharing a thought with someone else to interact directly with that person.

Third, ease of sharing: intuitively, even in one-on-one cases, I can communicate, agree, and disagree with someone with whom I have never before directly interacted. Meeting someone for the first time on the street, I can easily ask when the next bus is coming, agree that the buses should run more frequently, and disagree about whether this stop is served by a certain bus line. So, intuitively, sharing thoughts does not require more than minimal prior direct interaction between the parties involved, nor does it require much effort from either party in many cases.

Fourth, sharing in spite of differences: intuitively, two people can communicate, agree, and disagree even when they are very different from one another in many ways. I can meaningfully discuss buses with someone who does not believe all the same things about buses as I do, who is not inclined to judge all the same things to be buses as I am, who does not use buses in the same way I do, who does not bring the same image to mind on hearing the word "bus" as I do, who did not grow up talking about buses in the same language as I did, and who is different from me in relating to buses in other ways as well. So, intuitively, sharing thoughts does not require a great deal of similarity between the parties involved.

Last, but perhaps most important, sharing thanks to social interaction: intuitively, there are some cases in which communication, agreement, and disagreement enable sharing thoughts by allowing one party to have thoughts they couldn't have before. If I arrive in some new city that enjoys a mode of public transportation of which I was previously unaware and which I do not even come close to understanding—perhaps which I do not even realize is a mode of public transportation—I might be unable to have any thoughts about it until I engage some local in conversation about the city. Once the local mentions the mode of public transportation, I might gain the ability to discuss it with her further, agreeing with some of the things she says about it and disagreeing with others, even if I only barely grasp what she is talking about. In cases like this, intuitively, one gains the ability to share thoughts

[17]Cases like this raise important questions about indeterminacy. It is certainly plausible that "successfully communicates with" is a vague predicate, so one can fairly easily imagine sorites-type paradoxes concerning long chains of communication. This might seem to put some pressure on the idea that communication requires transitive sharing of thoughts. However, if it does then similar pressure could be put on the idea that predicates like "is exactly as tall as" are transitive. Such predicates are clearly transitive in an important sense—take "is almost as tall as" as an intransitive case for comparison—even if indeterminacy makes it difficult to quickly spell out what that sense is. I will discuss the indeterminacy of sharing thoughts and its relation to transitivity in more detail in chapter 5.

with another simply by interacting with them in the right way.

1.3.2 Robust publicity involves thoughts being made available by social context

Robust publicity, then, is meant to capture all these requirements and more, and so in itself as a desideratum it is stronger than minimal publicity. Minimal publicity simply involves actual widespread thought sharing, but robust publicity involves that and a bit more:

transitive sharing: Sharing thoughts is transitive, so that if A shares a thought with B, who shares the same thought with C, then A shares that thought with C.

sharing among groups: It is common for one thought to be shared by many individuals; thoughts are not only shared between pairs of individuals.

easy sharing: For many pairs of individuals, sharing thoughts is as easy as starting a conversation, with neither individual needing to put in much time or effort in order to be able to think something the other also thinks.

sharing despite differences: Sharing thoughts does not require a great deal of similarity between individuals.

sharing thanks to social interaction: Sharing thoughts is sometimes enabled *by* social interaction.

Most of chapter 5 of this work will be devoted to developing an account of the publicity of thoughts that is meant to do justice to the intuitions behind these claims and is also compatible with the assumptions presented above (that thoughts are individuated at least according to their truth conditions and that sharing thoughts requires sharing concepts). Because its development will involve several complicated arguments—most importantly the argument in chapters 2 through 4 about why certain ways of accounting for publicity are *not* compatible with those assumptions—I will not now present any more of the motivations for the account. Instead, I will just try to give a sense of the picture of the world that emerges from the account—a sense of what is really involved in robust publicity.

On the account I will develop, thoughts are public in that individuals within the right kind of community will be able to think certain thoughts because those thoughts are available to members of that community to be thought in a way in which they might not be available to members of other communities to be thought. Which thoughts an individual is able to think—which concepts an individual possesses—depends on the communities of which the individual is a member, among other things. I must, of course, say a bit more than this by way of a general description of publicity on the account, but first I think it might be helpful to present an example—an example to which I will return several times in the following chapters. What follows is robust publicity in action.

Imagine that A and B are both generally fairly competent users of English. A, being a competent English user, applies the word "cat" to cats of all kinds and is disposed, when asked the right questions, to say things like "Many cats meow," "Many cats have fur," "Cats are mammals," *et cetera*. A also possesses a concept, CAT, which A uses analogously to "cat," applying the concept to cats of all kinds and using it in the appropriate inferences. (We will ignore for now cases in which someone uses a word or concept in a way that they might, upon further reflection, consider to be a mistake.)

However, B is somewhat less competent with English, at least when it comes to "cat," since B only applies the word to furry cats. B is disposed, when asked the right questions, to say things like "Many cats meow," "Cats are mammals," *et cetera*, but B also says things like "All cats have fur," and when B encounters a naturally furless cat B is disposed to say things like "That isn't a cat. All cats have fur, and that animal does not have fur, so it must not be a cat." B also possesses a concept which B uses analogously to "cat," only applying the concept to furry cats and only using it in the corresponding inferences. And B does *not* possess any concept that B uses in the way A uses CAT, applying the concept to furry and furless cats alike.

I think the most natural thing to say about this case is that B possesses the concept CAT but has the false belief that all cats have fur—or perhaps is confused in some deeper way. And, indeed, this is the verdict that robust publicity demands, since B is intuitively capable of using CAT to, for instance, agree with A that a certain furry cat is a cat and disagree with A about whether a certain furless cat is a cat. But as I will argue in chapters 2 and 3, this verdict will not be available to many popular theories of thoughts that only determine which thoughts B has by taking into account how B is and how B relates to wider world without specially taking into account B's social context.

In order to give the right verdict in this and similar cases, and so to allow for robust publicity, a theorist must deliberately make room for publicity in her theories of thoughts by taking into account how B relates to the communities of which B is a member and how the other members of those communities are, as well as how B is *et cetera*. More specifically, as I will argue in chapter 5, a reasonable theory can only make room for robust publicity by taking into account how B interacts with B's fellow community members using public language.

On the account I will develop, the story of how B is able to think about cats has two parts. First, the thoughts that B expresses using the word "cat" are connected in a certain way, via public language, to certain thoughts of B's fellow community members. It is because of this that B is able to share thoughts with the other members of B's community. The question of which thoughts those are, though—the question of whether they are thoughts about cats, say—is not answered in this first part of the story.

Second, many of those thoughts involve a concept that is used in much the same way as A uses the concept CAT. It is (very roughly) because of this that those thoughts are thoughts about cats (rather than, say, thoughts about furry cats). Thus, B's position within the community makes the concept CAT available to B in a way that it would not be if B was otherwise the same but lived alone on a desert island or in a very different community.

This account, then, gives us a picture of concepts as entities that have social features more basically than they have at least some of their semantic features. If we pick out a concept other than by its semantic features, say as 'the concept B expressed by uttering "cat"' (if this can be a way of picking out a concept), and we then try to use the account to determine the semantic value of that concept, we will first have to determine the extent of the community of thinkers who share the concept. Once we have determined the extent of the relevant community, say the people capable of using the word "cat," we can go on to figure out the semantic value of the concept by looking at how everyone in the community uses it, say by looking at how everyone uses whatever concept they express by uttering "cat."

So concepts will turn out to be shared among the members of certain kinds of linguistic communities, with each concept determining the boundaries of one such community (though with a great deal of overlap among these communities). These communities will often be quite broad—for example, the people capable of using the word "cat," which is surely a very broad community, will turn out to be one—and all of the members of a community who share a concept will determine the semantic value of that concept together.

This picture is very sketchy now, but once the details are filled in it should be fairly clear how concepts will meet the requirements of robust publicity. And once the barriers to the publicity of concepts in which I am interested are removed, analogous barriers to the publicity of thoughts will also be gone.

This robust-publicity-involving picture might seem radical, but careful attention will show that it accords with many of our intuitions about particular cases, including cases that should already be familiar from various discussions in the literature[18]. Additionally, although this picture is incompatible with some popular theories of thoughts if those theories are taken to be exhaustive, the theory supporting this picture that I will develop in chapter 5 of this work can be seen as building upon, rather than replacing, current theories—or it could even be seen as just a detailed working through some of the implications of some already-proposed theories. At the end of this section, in subsection 1.3.4, I will compare robust publicity to a similar (or simply less specific) phenomena discussed by other authors, partly so that the philosophical ancestry of my project will be more clear. But before I do that, in the next subsection I wish to quickly discuss some of the theoretical costs that do and don't come along with this picture.

1.3.3 Robust publicity is not as problematic as it might at first seem

I have just given a preview of the account I will be developing over the course of this work, in order to give some sense of the direction the discussion is heading. However, the preview

[18]Likely the most famous case which motivates this kind of picture—of concepts as importantly social entities—is Tyler Burge's (1979) arthritis case. It seems to me that the only reasonable way to accommodate the intuitions Burge's case brings out is with a picture like this. Burge's case is similar in structure to the CAT case I presented above, though some of the details differ in ways that will be important to discussions in later chapters.

necessarily fell short of the full account in several ways that might seem to give readers cause for concern. In particular, I expect some readers are already wondering whether accepting the account involves taking on certain theoretical commitments that they find objectionable. Though I doubt I can put all readers at ease, I would like to spend this subsection briefly addressing several potentially problematic features that the account might seem likely to have. In doing so, I hope to save readers the suspense and distraction that come with keeping these potentially problematic features in mind until a later chapter when the relevant details of the account are made clear.

First, I should make clear that I do not intend the account I give of the semantic features of thoughts and concepts to be exhaustive: I wish to leave room for at least two ways in which these entities can fall outside the scope of the account I will develop. One way is that a person could have thoughts and concepts that are not related to the thoughts or concepts of other people in the way that 'public' thoughts and concepts are. For example, someone might have an idiosyncratic concept that they do not (or do not yet) express in any public language, say of a specific smell or of some newly-discovered subspecies. The account must leave room for such concepts and for their privacy, and it does because the only special room it makes for concepts to be shared is via public language. If someone has a concept that has no special connection to public language, the account has nothing special to say about that concept. I will say more about this in chapter 5.

The other way that thoughts and concepts might fall outside the scope of my account is that they might have more semantic features than those my account covers. For example, there might be some theoretical uses for which it would make more sense to determine the semantic values of the concepts an individual possesses based entirely on the way that individual alone uses those concepts, without taking the individual's social environment specially into account. In motivating my interest in publicity, I only want to claim that if thoughts are meant to be theoretical entities that are commonly the objects of communication, agreement, and disagreement, then thoughts must be importantly public. If instead thoughts are, say, only given the theoretical task of allowing precise prediction and explanation of individual behavior, then they might have truth conditions of a very different kind than those I am interested in here. As far as I can tell, there is no incoherence in distinguishing between theoretical roles that thoughts as truth-apt entities might play, whether this involves thinking of thoughts as having multiple kinds of truth conditions or thinking of an individual as having multiple kinds of thoughts. I will discuss the importance of these different theoretical roles at greater length in chapters 4 and 5.

I should also make clear that although I will offer an account of the publicity of thoughts that importantly involves public languages, I will not try to reduce the publicity of thoughts to the publicity of—say—public-language sentences or utterances. To do so would simply raise the question of how (apart from nominally) these public-language entities are themselves public, and I see little hope of answering that question while taking the publicity of these public-language entities to be more basic than the publicity of thoughts. Instead, the most promising strategy when it comes to accounting for the publicity of thoughts or of public languages will be to account for both kinds of publicity together, taking neither to be more

metaphysically or explanatorily basic than the other, and this is the strategy that I will pursue in chapter 5.

While on the subject of reduction, I might also say that although on my account many thoughts and concepts will have social features at a very basic level, the account will not require that these features are irreducibly social. For example, although linguistic communities will be important to concept individuation, they will not be important as irreducibly social entities, but instead as collections of interconnected thinking individuals. So on my account, facts about what an individual is thinking will importantly depend on facts about how the people around her are, but they will not depend on any further facts about how her community itself is. I will discuss this further in chapter 5, where I will also briefly discuss the alternative option of allowing irreducibly social entities into my account.

Lastly, I should go ahead and say that because linguistic communities (surely) admit of a great deal of indeterminacy, and my account ties the individuation and semantics of thoughts and concepts to linguistic communities, on my account the individuation and semantics of thoughts and concepts will admit of a great deal of indeterminacy. On my account there will often be no clear answer—or there will be many somewhat clear but inconsistent answers—to questions about what an individual is thinking, at least in part because of similar difficulties for questions about what communities the individual is a member of.

However, although indeterminacy concerning thoughts might seem to be a disappointing result, it should hardly be surprising, at least if thoughts are the kind of thing we can usefully talk about as the objects of communication, agreement, and disagreement among different individuals. Our minds and environments are so fantastically complex and subtly varied that it would be a great surprise if we could find any kind of common ground or points of useful comparison as rich as thoughts except at very high levels of abstraction, and abstracting to high levels often leads to indeterminacy. In chapter 2 I will focus at length on the difficulty of usefully abstracting away from the many and subtle variations among human minds, and I hope that by the end of that discussion some indeterminacy will seem like a small price to pay to overcome this difficulty.

1.3.4 Phenomena like robust publicity are not new as a topic of discussion

Before I end this introductory chapter, I would like to make one last attempt to clarify the problem I am interested in and the kind of solution I am seeking by briefly discussing Tyler Burge's treatment of a similar problem. In chapter 3, I will also discuss the treatments of similar problems offered by Jerry Fodor and Christopher Peacocke. One thing these treatments have in common is that they all involve acknowledgment of a phenomenon like robust publicity: all three authors hold that which thoughts a person is capable of having importantly depends on the person's social environment, even if they do not go so far as to claim, as I do, that the problem can be solved by acknowledging that concepts are importantly social entities.

1.3.4.1 Burge and arthritis

In "Individualism and the Mental," Burge brings out the intuitive appeal of something like robust publicity with a very thorough discussion of several thought experiments. Most famously, Burge gives two cases in which, intuitively, two individuals have different thoughts even though they differ only in their social environment. Burge presents the first case by describing a fairly ordinary individual:

> A given person has a large number of attitudes commonly attributed with content clauses containing 'arthritis' in oblique occurrence. For example, he thinks (correctly) that he has had arthritis for years; that his arthritis in his wrists and fingers is more painful than his arthritis in his ankles, that it is better to have arthritis than cancer of the liver, that stiffening joints is a symptom of arthritis, that certain sorts of aches are characteristic of arthritis, that there are various kinds of arthritis, and so forth. In short, he has a wide range of such attitudes. In addition to these unsurprising attitudes, he thinks falsely that he has developed arthritis in the thigh. (1979, 77)

This individual's thought is false because arthritis is an ailment specific to the joints: it isn't the kind of thing that can occur in the thigh. The individual simply makes the mistake of thinking that arthritis can occur in the thigh. And surely—the thinking goes—it is possible for someone to be mistaken about matters like this; surely we would describe many cases as involving mistakes of this kind, both cases we encounter from the outside and cases we find ourselves involved in, when we either correct someone else's mistake or accept someone else's correction of our own mistake.

Burge goes on to present the second case, featuring another individual:

> Precisely the same things (non-intentionally described) happen to him. He has the same physiological history, the same diseases, the same internal physical occurrences. He goes through the same motions, engages in the same behavior, has the same sensory intake (physiologically described). His dispositions to respond to stimuli are explained in physical theory as the effects of the same proximate causes. (1979, 77-8)

The two individuals are exactly alike in nearly all respects. The key difference is that in the broader community of which the first individual is a member, "arthritis" is standardly only applied to ailments of the joints, whereas in the broader community of which the second individual is a member, "arthritis" is standardly applied to ailments of the bones generally, so that it is perfectly normal to say things like "I have arthritis in my thigh" when one has certain symptoms.

The intuitive upshot of this merely social difference between the two cases is meant to be that when the first individual says "I have arthritis in my thigh" he expresses the false thought that he has arthritis in his thigh, but that when the second individual says "I have arthritis in my thigh" he expresses a true thought that is not about arthritis at all. After

all—the thinking goes—how could the second individual's thought be about arthritis when arthritis is an ailment specific to the joints, and it is the practice both of the individual and of the experts in his community to apply "arthritis" to ailments of the bones generally? The only connection between the second individual's thought and arthritis seems to be that the thought is expressed with the sound "I have arthritis in my thigh," which is the same sound the first individual uses to express a thought about arthritis, and this sameness of sound alone surely cannot be relevant. The lesson Burge would have us take, then, is that what an individual thinks depends importantly on the communities of which that individual is a member.

I am, of course, very sympathetic to this lesson. It is a central feature of robust publicity that communities are important in this way, and my main goal in this project is to motivate and account for robust publicity. I am also sympathetic to the motivations Burge presents, which mostly come down to a desire that our theories should do justice to our ordinary, inescapable ways of talking about particular cases. Insofar as my project is directly relevant to Burge's work, I am most inclined to see my project as supplementary.

One respect in which my project supplements Burge's work is that I offer additional motivation—or at least differently-developed motivation—for what I take to be Burge's main conclusion. Burge concludes that what an individual thinks depends importantly on that individual's social environment, largely on the basis of intuitions about particular cases. I wish to argue, along more theoretical lines, that if we are to hold on to anything like our ordinary picture of ourselves as thinking social creatures, our theories must have what an individual thinks depending importantly on that individual's social environment.[19]

The other respect in which my project supplements Burge's work is that I wish to show how a detailed theory of the semantics of thoughts might accommodate something like Burge's conclusion. And, of course, before doing this I will show how current theories fall short of this goal.

1.4 What comes next

In the following two chapters, 2 and 3, I will develop an argument that a wide class of theories of concepts are incompatible with the publicity of concepts. Given the assumption I made in this chapter that sharing thoughts requires sharing concepts, it will follow from this conclusion that these theories of concepts are incompatible with the publicity of thoughts.

I will call these problematic theories 'individualistic' theories of concepts, because what unites them is that they take questions of concept identity—say, the question of which concepts someone possesses—to be individualistic facts that do not depend on social questions in any special way. And I will contrast individualistic theories of concepts with what I will call 'social' theories of concepts. The distinguishing feature of social theories is that, like the

[19]Burge does offer some motivations along these more theoretical lines, but I wish to develop the argument in a bit more detail.

account I briefly outlined in subsection 1.3.2, they take questions of concept sharing to be prior to questions of concept identity.

Although I have already in this chapter given some reasons to accept the publicity of concepts as a desideratum, in chapter 4 I will argue at greater length that there is a need for theories of concepts that are compatible with—and that offer an explanation of—the publicity of concepts. Individualistic theories of concepts thus come with a cost, given that they are incompatible with the publicity of concepts.

Finally, in chapter 5 I will develop a social theory of concepts and show how it offers an explanation of the publicity of concepts. This theory will thus come with a benefit where individualistic theories come with a cost, and so the conclusion of this work will be that in order to hang on to the publicity of concepts and so hang on to the publicity of thoughts, theorists should adopt social theories of concepts.

Chapter 2

The problem of publicity

As I explained in the last chapter, in this work I will motivate a social theory of concepts. I am contrasting what I call social theories of concepts with what I call individualistic theories of concepts, the difference being in the order in which the theories answer the question of who shares a particular concept and the question of which concept that concept is, which typically includes the question of what that concept's semantic value is. Social theories first take up the question of sharing and only then proceed to the questions of identification and semantic value, while individualistic theories first take up the questions of identification and semantic value and only then proceed to the question of sharing.

Another way to put the difference involves comparing the priority of conceptual features: Social theories—the theories I am arguing for—take the thinkers who share a concept to be a more basic feature of that concept than that concept's semantic value. On the other hand, individualistic theories—the theories I am arguing against—take a concept's semantic value to be a more basic feature of that concept than the thinkers who share that concept.

I am trying to motivate social theories of concepts by arguing both that an important desideratum on theories of concepts is being able to accommodate the publicity of concepts and that individualistic theories cannot accommodate this publicity. My discussions of the importance of publicity and my development of a social theory of concepts can be found elsewhere in this work. My goal in this chapter and the next is simply to argue that individualistic theories of concepts cannot accommodate the publicity of concepts, even in a fairly minimal form. For the purposes of these chapters, I will be assuming that accommodating the publicity of concepts is an important desideratum on theories of concepts, and I will not be interested in developing a detailed social theory of concepts. Instead, I will focus entirely on arguing for the negative claim that individualistic theories of concepts cannot accommodate the publicity of concepts.

I will sketch the general form of this argument in section 2.1, and then in the remaining sections I will discuss each of the argument's premises in turn. However, without being applied to a specific theory of concepts the argument is more of a serious worry than the clearly devastating objection it is meant to be. So, in chapter 3 I will apply the argument to two specific well-worked-out individualistic theories of concepts that I hope will serve as

representative examples of two popular families of theories of concepts: Fodor's theory, which individuates concepts at least according to certain of their causal relations; and Peacocke's theory, which individuates concepts at least according to certain of their roles in something like inference. Before I get to the argument itself, though, I should make two quick points to clarify its aims.

First, for the purposes of this argument, the 'publicity' of concepts need involve nothing more than many people actually sharing many concepts—in the terms introduced in the last chapter, I will here be concerned with 'minimal' publicity. Since the individualistic theories with which I am concerned offer possession and individuation conditions for concepts, the question of whether two people share a concept is straightforward: two people, A and B, share a concept just in case one of the concepts A possesses is the same concept as one of the concepts B possesses. The question of whether a particular concept is 'public,' then, is also fairly straightforward: a concept C is public just in case many people actually possess C.

What I will be arguing is that individualistic theories of concepts have the unfortunate entailment that many concepts are not shared by more than a few people, even though we would ordinarily think of these concepts—concepts like CAT, RED, HOUSE, and KNOWS—as typical examples of public concepts. On these theories, the class of concepts that are widely shared will be extremely limited, so limited that if we were to adopt these theories we would have to radically revise our conception of ourselves as social, thinking creatures.

And second, in this argument I will only be concerned with challenges to publicity relating to the semantic values of concepts. So, for example, even though there might be grounds to worry that Fodor's theory is incompatible with the publicity of concepts because Fodor individuates concepts in part according to their modes of presentation, which might seem likely to vary from one individual to the next, I will not argue this. Instead, I will argue that individualistic theories like Fodor's and Peacocke's are incompatible with the publicity of concepts because for nearly every case that we might care about in which A possesses a certain concept and seems to share it with B, these theories will imply that A's concept has a different semantic value from every one of B's concepts. Given the assumption (discussed in the last chapter) that concepts are individuated at least according to their semantic values—an assumption that Fodor and Peacocke both share[1]—this will imply that A does not share the concept with B.

With those points made, I will now proceed to the argument.

[1]For examples, see Fodor (1998, 15) and Peacocke (1995, 16). It should hardly be controversial that concepts are individuated at least according to their semantic values, at least in the situations which will be relevant to my arguments.

2.1 Why individualistic theories cannot account for publicity

Throughout this chapter, my discussion will focus on general concepts like CAT, WATER, and RED—concepts which we use to denote properties or classes of objects in our thought, rather than, say, individuals, relations, or events. This is for the sake of clarity. The extension of my argument to cover other kinds of concepts should be fairly straightforward.

Also for the sake of clarity, I will treat these general concepts as each denoting some property (theoretically equivalent here to the class of objects having that property) as its semantic value. Thus, if two thinkers denote two different properties in their thoughts, they are using different concepts. This may well be a strong simplification, since some general concepts might seem to denote different properties in different contexts of use. For example, CAT and WATER might have 'hidden' indexical elements[2], denoting whatever deep property explains the presence of certain surface properties in the environment in which the concept was acquired. Or these concepts might have as their semantic values vastly more complex functions from contexts, including both contexts of use and contexts of evaluation, to properties[3].

I do not wish to take a stand on these issues of context dependence, so I will focus on cases in which different thinkers are in relevantly similar contexts. The concepts I will be discussing might simply relate individuals to properties, or they might relate individuals to different properties in different contexts. Either way, in the cases I will be discussing, if two thinkers denote two different properties in their thoughts, they must be using different concepts. For example, suppose that WATER denotes properties indexed to planets, such that a thinker on Earth and a thinker on Twin Earth could both use that concept to think about the clear, nourishing liquid filling their lakes and streams. Then the fact that an Earthling and her twin have thoughts concerning different properties—one *being composed of* H_2O and the other *being composed of* XYZ—does not show that they have different concepts. But if WATER is indexical only in this way, and we instead consider two Earthlings thinking thoughts they would express using the word "water," then if their thoughts concern different properties, that would be enough to show that they have different concepts. This is because in this latter case the planet-indexicality of WATER is not a factor, because all the action takes place on the same planet. I will only focus on cases like this, in which there is no relevant variation of context between the two individuals concerned.

However, I will not actually be focused on any WATER cases at all. Among general concepts like CAT, WATER, and RED, I will further focus on concepts which denote relatively complex properties, like CAT and RED, rather than on concepts which might plausibly denote quite simple properties, like WATER. And this focus is not just motivated by presentational concerns: as I discuss below, my argument will only apply to concepts with relatively

[2]This relatively simple kind of hidden indexicality was famously proposed by Putnam (1973).

[3]This context dependence of great—if not infinite—complexity is motivated by cases like those discussed by Travis (1985).

complex denotations. And among complex general concepts like CAT and RED, my discussion will be even further focused on 'natural kind' concepts like CAT, since they might seem to be the easiest targets when it comes to accounting for publicity—if individualistic theories of concepts cannot account for the publicity of relatively natural concepts like CAT, there is little hope that they will be able to account for the publicity of less natural concepts like RED.

The rough structure of my argument is this:

P1: The concept CAT has an arbitrary denotation, even if it is somehow vague.

P2: If a concept has an arbitrary denotation, then if a theory of concepts does not explain how many people possess concepts with that denotation, on that theory we should not expect many people to share that concept.

P3: Individualistic theories of concepts cannot explain how many people possess CAT.

P4: Other complex general concepts are like CAT with respect to P1 and P3.

C: Therefore, on individualistic theories of concepts we should not expect many people to share complex general concepts.

This is the argument in its general form, but I do not expect it to be persuasive—even if it is comprehensible—without further discussion of some of the premises, especially premise 1. So, over the remaining sections in this chapter, I will discuss each of the premises in turn, so that I might explain the premises that need explanation and justify the premises that need justification.

2.2 There is arbitrariness in the denotation of CAT, even if it is vague

This first premise will probably be the least clear on first reading, and once it is made clear it will probably be the most controversial. By way of making the premise more clear, I will first say a bit about what I mean when I talk about arbitrariness in general, and I will then say a bit about what I mean when I talk about the arbitrariness of the denotation of CAT in particular. Once the premise has—I hope—been made fairly clear, I will try to justify it by saying a bit about why the denotation of CAT is arbitrary and by addressing a couple of potential objections. The first objection has to do with the 'naturalness' of the denotation of CAT and whether that can ground a strategy for resolving arbitrariness: I will show that it cannot, because the denotation is arbitrary even with respect to naturalness. The second objection has to do with the vagueness of CAT (or at least of "cat") and whether that can ground a strategy for resolving arbitrariness: I will show that it cannot, because the arbitrariness and vagueness are distinct phenomena, and the most plausible way the two are related is that vagueness is grounded in arbitrariness, rather than the other way around.

2.2.1 Arbitrariness

To say that something is arbitrary is to say that it could just as well have been different or have gone differently. For example, it is in some sense arbitrary that in the United States we drive on the right side of the road: we could just as well drive on the left side of the road, as long as we all followed this convention and there were corresponding reflections in related conventions, such as the convention governing which side of a car the steering wheel goes on.

In claiming that something like the denotation of CAT is arbitrary, then, I am claiming that it could just as well have been different from what it is. This claim raises several questions, among them the questions "Different how?" and "Just as well by what standard?" In each case of arbitrariness, there is some group of possibilities or options and some group of standards or purposes which do not distinguish among those possibilities or options. These are the groups in virtue of which the case is a case of arbitrariness, and the relevance of the arbitrariness is proportional to the relevance of these groups to the case. In a case in which an agent has to make a choice which is truly unqualifiedly arbitrary for that agent, the group of options must be the group of all good options open to that agent in making that choice, and the group of purposes which do not distinguish among those options must be the group of all the agent's purposes that are relevant to that choice.

For example, our historical collective choice as drivers in the United States to drive on the right side of the road is arbitrary because our two good options—driving on the right and driving on the left—are not obviously different from one another with respect to most of our purposes as drivers—safety, fast arrival, carrying cargo and passengers as necessary, and so on. The choice is likely not fully arbitrary, because we drivers are not entirely symmetrical, and our asymmetry—perhaps most relevantly, the prevalence among us of those who are right-hand-dominant—is likely responsible for very subtle differences between our options of driving on the right and driving on the left when it comes to at least some of our purposes as drivers, such as safety. But the choice is still very much an arbitrary one.

As I will discuss in more detail in the next subsection, I am concerned with one particular upshot of arbitrariness in cases in which multiple people individually make arbitrary choices or otherwise individually arbitrarily select[4] one possibility out of many. That upshot is that in such cases there is no expectation of coincidence: without some other reason to think that they will, we should not expect people's selections to match. For example, given the relative arbitrariness of which side of the road a country's drivers drive on, if a driver from the United States is traveling to some foreign country, unless she has some special knowledge about how the residents of that country drive, she should not expect the drivers in that country to drive on the right side of the road.

[4]I mean "select" here in a non-agential sense: the cases I am interested in include cases in which multiple people are individually responsible for determining which possibility out of some group of possibilities becomes actual, even though the people do not deliberately choose that possibility. For example, in a case in which multiple people walk through the same door, it is arbitrary which foot is in front when each person steps through the door, and each person is responsible for determining which of his or her feet is in front when he or she steps through the door, even though none of the people makes a choice about it.

2.2.2 Standards relevant to the denotation of CAT

Given what I take arbitrariness to amount to, my claim that the denotation of CAT is arbitrary should immediately prompt two questions. One is the question of what the relevant standards governing CAT denotations are, and the other is the question of what the possibilities for CAT denotations among which these standards do not distinguish are. If I can establish that all the relevant standards do not distinguish among all the relevant possibilities, then I will have established—to my own satisfaction, at least—that the denotation of CAT is arbitrary.

There are, of course, many standards which are potentially relevant to the denotation of CAT. Or, at least, there are many standards which are not obviously irrelevant to the denotation of CAT. So I will not try to give an exhaustive discussion of all of these standards and show how each standard in turn is not sufficiently discriminating to distinguish among many possible denotations of CAT. Instead, I will make some general remarks about two kinds of standards which might seem relevant, and then I will leave it up to the reader to decide whether any particular relevant standards can distinguish among the various possible denotations I will discuss in the next subsection.

One kind of standard which is surely in some way relevant to the denotation of CAT is the kind of standard that semantic theories of concepts give us: standards that, if correct, govern what the denotation of CAT actually turns out to be. And if I were making a certain kind of argument about the denotation of CAT, I would indeed have to take up these standards as relevant to what I am doing and go on to show how they cannot distinguish among many possible denotations.

However, these standards purportedly governing the actual denotation of CAT would only be relevant at this point in my argument if I were trying to motivate the premise that the denotation of CAT is indeterminate, and that claim is importantly different from the claim I am making here, which is the claim that the denotation of CAT is arbitrary. In the present context, the difference between indeterminacy and arbitrariness might be put like this: in a case of indeterminacy, multiple possibilities are actual, whereas in a case of arbitrariness, multiple possibilities are the best—or are good enough, in cases of less than full arbitrariness. My argument can succeed in establishing my conclusion whether the denotation of CAT is determinate or indeterminate, so I am not now concerned with indeterminacy.[5] I will have

[5]Another kind of case in which arbitrariness is important and indeterminacy is not is a kind of case that game theorists discuss: coordination problems. These problems typically exhibit arbitrariness because they feature multiple players each of whom has multiple options, none of which is distinguished by the relevant standards as the best option. However, it is in the players' interest to somehow overcome this arbitrariness, because it is better if more players pick the same option and worse if more players pick different options from one another. Arbitrariness is thus of great importance to these cases, but indeterminacy is not because these cases need not exhibit any indeterminacy: it might be fully determinate which option each player ends up picking without the case becoming any less interesting.

Which side of the road to drive on is a classic example of a coordination problem, since it exhibits the right kind of arbitrariness, as I discussed above, and it is in the interest of the players that they all select the same side, whatever side that might be. But these facts about the goodness of the options can stand

something to say about it in a later subsubsection, though, when I discuss how vagueness is and is not relevant to the arbitrariness of the denotation of CAT.

The relevant standards, then, are those which distinguish good denotations of CAT from bad denotations of CAT, rather than those which distinguish the actual denotation or denotations of CAT from everything else. Since the upshot of arbitrariness with which I am concerned here is the lack of expectation of coincidence, if my argument is to succeed then these standards must include everything that might plausibly motivate thinkers to acquire a concept with one of the possible denotations of CAT[6] rather than another—just as in the driving case discussed above, if it had been a case of unqualified arbitrariness the purposes in virtue of which the decision was arbitrary would have included all purposes that nations of drivers might have in collectively deciding which side of the road to drive on.

And what standards are in play, potentially motivating, when a thinker acquires a concept like CAT (or, if this is a separate event, when a thinker with a concept like CAT fixes the denotation of the concept)? The most obvious standards are pragmatic: some denotations make for a more useful concept than others. For example, if instead of the concept CAT, which we use to think about cats, we had the concept CAT-OR-RED, which we used to think about things that are cats or colored red, our lives would be more difficult in various ways. As pet owners, we wouldn't be able to sort ourselves quite so neatly into dog people, cat people, and others; the best we could do instead would be to sort ourselves into the relatively easily understood category of dog people, the small and strange category of people who like cats and red things, and others. As aficionados of funny internet videos, we would have a harder time effectively categorizing those videos for archival and sharing. And as zoologists, without coming up with an entirely new concept we would not even be able to take up cats as a subject of study.

It seems to me to be open to debate whether there are standards distinguishing good denotations from bad denotations of concepts like CAT other than pragmatic standards. One potential standard has to do with how 'natural' the denotation is—how well it does at carving nature at its joints—and I will discuss this standard in relation to my argument in the next subsection. And there may well be other standards which are potentially relevant here. However, even if there are other relevant standards, I do not think they will ground any strong objections to my argument, as I will now explain.

The premise of my argument that is the current topic of discussion is the premise that the denotation of CAT is arbitrary, and the subconclusion that this premise is meant to help

whether or not there is a fact of the matter about which option each player takes once the game is run: the case exhibits arbitrariness whether or not it exhibits indeterminacy.

[6]It is somewhat strange to talk about possible denotations of CAT, since it is fairly reasonable to think that if the concept CAT has any property necessarily, then it has its denotation necessarily, and so the possible denotations of CAT just are the actual denotations of CAT (whatever those are). But the kind of possibility that is relevant here is epistemic, rather than metaphysical: the possible denotations of CAT that are relevant are those properties or sets of object (whatever denotations in general are) that we as theorists should be considering when we are trying to answer the question of what the denotation of CAT is. So unless we have already definitively and correctly answered this question, we should allow that there are some possible denotations of CAT which are not actual denotations of CAT.

establish is the subconclusion that if a theory of concepts does not independently explain how many people acquire concepts with the denotation of CAT, then on that theory we should not expect many people to share CAT. So the potential objections that are relevant to this discussion are objections that purport to establish that the denotation of CAT is not arbitrary and so that we might expect many people to share CAT even if our theory of concepts does not give us some kind of independent explanation of this phenomenon.

Thus, if there are any standards distinguishing good denotations from bad denotations which might ground a strong objection to my argument, those standards must give us some reason to think that many people will acquire concepts with the denotation of CAT. This will only happen if we have reason to think that these standards are somehow motivating concept-acquiring thinkers, influencing which concepts these thinkers acquire—or, more precisely, influencing with which denotations these thinkers acquire concepts. It should be fairly clear that pragmatic standards are potentially motivating in this way, since the relative usefulness of concepts is something that thinkers might become aware of and thus take into account when acquiring concepts, though perhaps not deliberately. And if there are any other standards which are relevant to my argument here, they must also involve features of concepts which thinkers might become aware of and thus take into account when acquiring concepts, though—again—this need not be a deliberate process.

However, as should become clear when I discuss below some of the possibilities or candidates in virtue of which the denotation of CAT is arbitrary, many of these candidates differ only in ways that are undetectable to thinkers in ordinary contexts. The upshot of this is that any standards which distinguish among these denotations will do so based on features of which thinkers are ordinarily ignorant and about which thinkers do not ordinarily care, so it will not be plausible that many thinkers somehow use these features to coordinate their concept acquisition and thus all acquire concepts with the denotation of CAT.

In the next subsection, then, I will discuss candidate denotations of CAT, and my goal will be to show that there are many different candidates each of which is such that there are no candidates that are better in any way that we could expect thinkers to pick up on. In my discussion, I will focus on pragmatic standards—on how useful concepts with different denotations would be to us as pet owners, internet video aficionados, and zoologists—and on showing how there are no potentially motivating pragmatic differences among the candidates. (And in the subsection after that, I will focus on standards based on how 'natural' candidate denotations are.) However, the candidates I discuss will differ so subtly that differences among them would either go undetected or would, if detected, not be significant enough to ground anything like a preference for one denotation over another, so even if there are standards other than those having to do with pragmatics (and 'naturalness') that are potentially motivating when it comes to acquiring concepts, they will do no more to distinguish among the candidates I offer than will the standards I explicitly discuss.

2.2.3 Possible denotations of CAT

Having *cathood*, the property denoted by the concept CAT, might plausibly amount to having a certain very complex genetic property or to being descended from a certain population[7]. But which complex genetic property, or which population? Surely there will be many candidates for which specific property *cathood* amounts to, all good enough to be the denotation of CAT. This is because for almost any given complex genetic property which is itself a good candidate denotation, there will be many similar complex genetic properties differing only in relatively small ways which are also good candidates, which indeed are no worse as candidates than the given property. And similarly, for almost any population which is a good candidate, there will be many similar populations differing only in relatively small ways which are also good candidates, which indeed are no worse as candidates than the given population.

When I say that CAT has an arbitrary denotation, part of what I mean is that there is this proliferation of indistinguishably good candidate denotations. If we had some independent way to identify CAT as the concept under investigation, and we were asked to identify its denotation, we would be spoiled for choice—we would have no principled way to select among the many candidate denotations, because no reasonable criterion would distinguish one (or even a few) of the good candidates from all the others. This is, of course, a very strong claim, since there are many reasonable criteria by which we might try to distinguish the true denotation of CAT from among the many mere candidate denotations. But I think working through the details of some of the ways we might try to identify the denotation of CAT will help to make clear that no matter what reasonable criterion we try to apply to single out the best candidate, there will be many candidate denotations that all meet that criterion and yet are clearly distinct from one another.

For example, let us suppose that we have already determinately settled which currently living animals are cats, and that any set of genes possessed by all currently living cats and no other currently living animals makes a good candidate for the complex genetic property which *cathood* amounts to. There will be many such sets, because for most such sets the removal of one gene or the inclusion of one more gene possessed by all currently living cats will make more such sets. So there will be many good candidates. And although there might be no distinguishing among the candidates when it comes to the question of which currently living animals are cats, the different candidates will have different implications for which past, future, and merely possible animals count as cats.

Or let us suppose that we have already determinately settled which currently living animals are cats, and that any suitably temporally and geographically clustered population of animals, all of whom are ancestors of all currently living cats, makes a good candidate for the population descendancy from which *cathood* amounts to. There will be many such populations, because for most such populations the removal of one individual or the inclusion

[7]I assume that cats did not all descend from a single breeding pair, or at least that in developing our theories of concepts we cannot count on concepts like CAT denoting descendants of a single breeding pair or some similarly easily delineable class.

of one more individual, ancestor to all current cats, living nearby around the same time, will make another such population. So there will be many good candidates. And although there might be no distinguishing among the candidates when it comes to current or future cats, the different candidates will have different implications for which past and merely possible animals count as cats.

These examples were surely greatly oversimplified in several respects, but realistically complicating them will not greatly reduce the numbers of good candidates. For instance, there is no obviously correct way to determinately settle which currently living animals are cats, especially if cats as a domesticated species are to be distinguished from wildcats, but there being multiple good candidates for which currently living animals are cats does nothing to simplify the situation. If anything, it makes for more good candidate properties for the denotation of CAT.

Another way to bring out the fact that there are many indistinguishably good candidates for the denotation of CAT, starting from what standards of goodness might be in play rather than from what might distinguish different candidates from one another, is to ask why it is that we use CAT rather than a concept with a different denotation. Is there some special factor that distinguishes CAT from all its possible peers, other than the fact that it is the concept that we turned out to use? Given the complexity of CAT which the above discussion brought out, it is far from clear to me what this special factor could be.

If *cathood* does amount to having a certain very complex genetic property or to being descended from a certain population, then there will be many other similar properties which could just as well have been the denotation of CAT. That is, if in our community we use CAT, which denotes some determinate property *cathood*, then there is no clear reason to think that had we been slightly different we would not use a different concept, denoting some similar but subtly different property—CAT_1 denoting $\textit{cathood}_1$, CAT_2 denoting $\textit{cathood}_2$, or one of many others—a different concept which would serve our purposes as well as CAT does. Using CAT_1, CAT_2, or some other concept only subtly different from CAT would allow us to care for pets, share amusing videos on the internet, and even practice zoology just as well as we do using CAT.

Indeed, someone using one of these alternative concepts when everyone else uses CAT would likely go undetected, since the subtle differences between the alternative and CAT would not show up in behavior in typical situations. It would take some fairly intense conceptual analysis, involving perhaps some imaginary, merely possible cat hybrids or some very distant cat ancestors, to bring out a difference. But the differences would still be there, and on an individualistic theory of concepts they could make all the difference between sharing concepts and not sharing concepts.[8]

[8]Indeed, as I will make clear in this chapter, the picture of concept possession we get on individualistic theories is a picture in which everyone is using their own personal alternative concept instead of CAT, but these idiosyncracies only typically become clear during intense sessions of conceptual analysis, and otherwise they only manifest in a great deal of undetected miscommunication.

2.2.4 Naturalness and eligibility

One way to cast what I am doing in this chapter is as trying to make trouble for a certain family of semantic views in part by arguing that views in the family do not do enough to narrow the field of candidate semantic values for a given term. For example, I am arguing that CAT in some sense might have as its semantic value one of any number of properties and that—in part because there are so many good candidate semantic values out there—on individualistic theories of concepts it will simply be a matter of chance whether two people with concepts in the vicinity of CAT turn out to have concepts with the same semantic value.

Viewed this way, my objection to individualistic theories of concepts bears a certain similarity to an attack on interpretationist theories of semantics going back at least to Putnam (1980). The details of the attack and of the theories are not relevant here, except that Putnam showed that at least some theories in that family had a serious problem in that they left open which of a great many candidate semantic values a given term had. And since this upshot of Putnam's attack closely resembles one of the upshots of my attack, we might look to responses to Putnam's attack on behalf of interpretationists for potential responses to my attack.

The response I want to consider was most famously offered by Lewis (1983). Lewis suggested that the interpretationist might narrow the field of properties which are candidates for the semantic value of a given term by finding the property that was the most metaphysically 'natural' and claiming that that property was most 'eligible' to be the semantic value of the term. The notion of a property's being metaphysically natural is the same as or closely connected to the notions of the property's being metaphysically 'basic,' 'fundamental,' or 'primitive,' or of the property 'carving nature at its joints.' The basic idea is that some properties—perhaps those described in basic physics—are specially privileged, or at least most privileged, whereas other properties—perhaps those the instantiation of which merely supervenes on the instantiation of the properties described in basic physics—are not privileged, or at least are less privileged.[9]

We can see Lewis's response, as it would be applied to the case of the concept CAT in which I am currently interested, as involving two distinct parts. One part is an argument that for CAT and other concepts like it, only one of the properties which are otherwise good candidates for being the semantic value is natural in the relevant sense, or at least only one is most natural. The other part is an argument that if there are many candidate semantic values for a given concept, but one is most natural, then other things being equal that most natural candidate is the semantic value of the concept. I will not be discussing this latter part of the strategy, so I will just grant Lewis that more natural properties are more eligible to be semantic values.

[9]Naturalness and fundamentality are important and often-discussed notions in metaphysics, and giving a full account of either of them would be going beyond the scope of this work. These notions are important here, dialectically, simply because they offer the promise of some principled way to distinguish properties as good or bad candidates for the semantic value of a given concept, beyond looking to the practical significance of the properties.

The question, then, is whether there really is just one property among all the otherwise good candidates which is the most natural. The answer, of course, depends on what it is to be natural, and this varies from one theory to another. For the purposes of this argument, we can consider theories of naturalness as they fall into two categories. These categories are not exhaustive (or, as will be seen, even exclusive), but they do exhaust the ways that metaphysical naturalness or a similar notion might be relevant to the dialectic here.

Theories of the first kind posit a special, privileged class of primitive, basic, or perfectly natural properties, which are 'sparse,' rather than 'abundant,'[10] so someone wishing to deny the arbitrariness of CAT might claim that *cathood* is primitive but that primitive properties are so sparse that neither $cathood_1$, nor $cathood_2$, nor any of the other nearby properties are themselves primitive. If this were the case, *cathood* would be distinguished from all its neighbor properties, and so CAT would not be arbitrary. What grounds, then, could there be for claiming that *cathood* is primitive but $cathood_1$ is not?

At least at first glance, it would be very strange to claim that *cathood* is primitive, since typical examples of primitive properties are properties like *having a negative charge* or *massing 4 grams*, and *cathood* seems rather different from these. Even if there is no analysis of *cathood* which is clearly privileged over all other analyses, *cathood* certainly seems to admit of analysis—*cathood* might amount to a complex genetic property, or it might amount to a complex ancestral property, or it might amount to something else, but it certainly seems to amount to something other than just *cathood*. So even leaving $cathood_1$ aside, it would be difficult to justify the claim that *cathood* is primitive.

However, there are senses of 'primitive,' as well as background metaphysical theories, on which *cathood* is primitive, so we should not immediately reject the idea. For example, if every class of possible objects identifies the 'primitive' property which all the objects instantiate, and *cathood* is identified like this, then it will be primitive in the relevant sense. But bringing in this special class of distinguished properties still will not help with the arbitrariness of CAT, since primitive properties of this sort will be abundant enough that it will be easy to identify many primitive $cathood_1$s. It will not be helpful, then, to try to claim that *cathood* is primitive but that, because primitive properties are sparse, $cathood_1$ is not.

The other kind of theory which I wish to discuss here includes theories which hold that some properties are more natural or basic than others. Strictly speaking, this claim does not entail the claim which distinguished the first kind of theories—the claim that there is a class of perfectly natural or completely basic properties. However, one of the most straightforward ways to analyze the degree of naturalness of a property is in terms of the ease with which the property can be defined in terms of perfectly natural properties, so the two claims can come together.

This second family of theories is dialectically relevant because, although it might not be very plausible that *cathood* is primitive or a perfectly natural property, it is surely quite

[10]Lewis (1986) introduced the "sparse"/"abundant" terminology for this distinction in how 'thick on the ground' properties are. For an example of a theory on which universals, a certain kind of privileged property, are sparse, see Armstrong (2010).

plausible that it is more natural than many other properties. So someone wishing to deny that CAT is arbitrary might claim that *cathood* is more natural than $cathood_1$, $cathood_2$, and all the other candidate denotations. And if *cathood* were the most natural of the candidate denotations, it would be distinguished in a way that could resolve the arbitrariness of CAT.[11]

The problem here is that, as I emphasized above, for any plausible analysis of *cathood* there will be many ways to come up with a $cathood_1$ which is no less natural—by slightly altering the genetic requirements for possessing *cathood*, for example, or by identifying a slightly different ancestral population. In light of the number and complexity of the candidate *cathood* properties, there is no clear reason to expect to find a single candidate that is significantly more natural than every other candidate.

Of course, this depends on what naturalness of properties amounts to. I hope it is fairly intuitive at this point that whatever naturalness of properties amounts to, it is unlikely to single out one of the candidate properties from all the others. However, because there is not a single well worked out, widely accepted view of naturalness, it is difficult to give this intuition very solid theoretical backing.

Nonetheless, we might consider the question of how the naturalness of *cathood* and $cathood_1$ compare while adopting a view of naturalness that is well known even if it is not at all widely accepted, in order to get some idea of how the same question might be answered on other views of naturalness. So let us suppose that the degree of naturalness of a property amounts to the length of the shortest definition of that property in a logical language which only has terms for numbers and basic physical properties.[12] If *cathood* amounts to a genetic property it will, of course, have a very long definition in this language, but it might still be possible to compare the length of *cathood*'s definition to the length of $cathood_1$'s definition.

The shortest way to define these complex genetic properties in terms of numbers and basic physical properties will likely be to define them in terms of simple genetic properties, which are themselves defined in terms of chemical properties, which are finally defined in terms of basic physical properties. Since both *cathood* and $cathood_1$ will involve the same simple genetic properties, any differences in the lengths of their full definitions will come

[11]This is roughly the strategy followed by Lewis (1983) responding to Putnam (1980). Sider (2011) offers a more recent discussion of the degrees to which properties carve nature at its joints and their relevance to questions of taxonomy, though—like me—he does not claim that there will always be a maximally joint-carving candidate meaning for terms like "cat." This is in the context of a larger discussion of what it is for a question to be 'substantive,' in the sense that the answer to the question does not depend on the joints along which nature is carved:

> Except for questions that strain the boundaries of taxonomy (a relatively uncommon occurrence), special-science questions normally fall into one of the following two categories: i) each expression has a candidate meaning that carves far better than all other candidates; or ii) each expression has a range of candidates that carve far better than do other candidates not in the range, and the question's answer is insensitive to choices of candidates within these ranges. In either case, the question is substantive. (Sider, 2011, 48)

My position is that the question of what *cathood* is falls into the latter category.

[12]Lewis (1983) offers an analysis of the degree of naturalness of a given property that sticks roughly to these lines.

down to the lengths of their definitions in terms of simple genetic properties. *Cathood*, then, will be the most natural of all the candidate denotations if it has the shortest definition in terms of simple genetic properties.

However, we have no clear reason to expect there to be any definition of a candidate denotation which is shorter than all the others. For any good candidate which has a definition of a certain length, we should expect to be able to find another good candidate with a definition of the same length by simply altering the definition slightly. For example, to greatly simply things, let us suppose that there are 200 binary genetic properties relevant to *cathood*, and that one good candidate for the denotation of CAT is the property *having at least 100 of these 200 properties*. If that were the case, the slightly more lenient property *having at least 99 of these 200 properties* and the slightly stricter property *having at least 101 of these 200 properties* would probably also be good candidates, and their definitions in terms of numbers and basic physical properties will be equally long. This is, of course, a rather fanciful example, but we have no reason to expect the truth to be any different.

We simply cannot count on there being a single most natural good candidate denotation if *cathood* amounts to a genetic property and naturalness amounts to ease of definition in terms from basic physics. And—although I hope I do not have to discuss the possibility in as much detail—the same goes if *cathood* amounts to an ancestral property, since it would admit of structurally similar variations.

As I already mentioned, metaphysical naturalness and related notions are important and still a matter of debate, and I could not give a full picture here of their relevance to my argument. However, I think that in light of the complexities of *cathood* which I have outlined, it should be fairly clear that on any reasonable view, even if something like the naturalness of properties is relevant to their eligibility for being the semantic value of a concept, naturalness cannot resolve the arbitrariness of CAT.

2.2.5 Vagueness and arbitrariness

One reasonable response to the arbitrariness involved in selecting a single, well-delineated property as the denotation of CAT is to refuse to do so, instead taking CAT to vaguely denote one of a family of such properties or to denote a single vague property. However, this does not solve the problem of arbitrariness (if it is a problem). Though selecting a determinate denotation for concepts like HEAP might clearly involve arbitrariness in some unacceptable way, it would be strange to suggest that treating such concepts as having vague denotations leaves no room for arbitrariness—that HEAP, considered as a vague concept, has something specially distinguishing it from all other concepts with subtly different denotations, like COLLECTION SHAPED LIKE A HEAP BUT SLIGHTLY LARGER[13].

[13]As they are written here, there might seem to be an important distinction between the concepts HEAP and COLLECTION SHAPED LIKE A HEAP BUT SLIGHTLY LARGER: the latter is a complex concept, but the former is simple, or at least simpler. But, of course, there is no reason to think that all concepts with denotations which are only subtly different from the denotation of HEAP will be complex, or at least more complex than HEAP.

To see how there can still be arbitrariness to CAT, even considered as a vague concept, let us suppose that there are only determinate properties, and that it is indeterminate which one out of a family of (determinate) properties CAT denotes. Let us further suppose that 200 of the properties which are candidate denotations of CAT can be ordered along some single dimension so that on one side there are properties which are clearly bad candidates for the denotation and on the other side there are properties which are clearly good candidates for the denotation. So within the ordering of these candidate properties c_1, c_2, ..., c_{199}, c_{200}, c_1 is clearly a bad candidate, c_{200} is clearly a good candidate, and for any candidate, c_n, the next candidate, c_{n+1}, is if anything a better candidate, but there is no pair of adjacent candidates such that one is clearly bad and the next is clearly good.

Vague terms or concepts typically exhibit a structure like this. For example, imagine that we are trying to determine the denotation of the term "heap," and that to simplify things we are looking for the property of being a contiguous collections of grains of at least a certain number. We might at one point find ourselves considering the properties h_1, h_2, ..., h_{199}, h_{200}, where each property h_n inheres in all and only contiguous collection of grains numbering at least n. Counting single-grain collections as heaps seems far too lax, so h_1 is clearly not a good candidate denotation, and both counting 199-grain collections as non-heaps and counting 200-grain collections as heaps seem individually reasonable (though not reasonable taken together), so h_{200} is clearly a good candidate, or at least as good a candidate as any. And there is surely no pair of adjacent candidates such that one is clearly bad and the next is clearly good. This is precisely the structure of the candidate denotations for CAT in our supposition above.

Given the complexity of *cathood*, there should be many groups of properties and many dimensions which fit our supposition. For example, if *cathood* amounts to a genetic property, we might take 200 candidate properties which, starting from a clearly good candidate, include more and more genes which express themselves in more and more dog-like anatomies, so that the final candidate inheres in creatures far too dog-like to be considered cats. Or, if *cathood* amounts to an ancestral property, we might take 200 candidate properties which, starting from a clearly good candidate which identifies the relevant population—from which all currently living cats are descended—as living in a certain place at a certain time (say, 100,000 years ago), move forward in time in identifying the relevant population (by, say, 500 years per candidate), so that the final candidate identifies a population far too dispersed (or just too recent) to plausibly be considered the original population of cats.

Given such a group of 200 candidates, different theorists concerned with vagueness might describe the denotation of CAT differently. However, any reasonable description will still leave considerable room for arbitrariness. To see descriptions in line with three prominent families of theories of vagueness, and to see how they do not resolve the arbitrariness of CAT, let us consider 200 claims of the form "all objects with property c_n are cats," substituting the numbers 1 through 200 for n and assuming that "cat" and CAT share denotations.

An epistemicist[14] might claim that there is some i between 1 and 199 such that claims

[14]Williamson (1994) and Sorensen (2001) offer examples of epistemicist theories of vagueness.

1 through i are false and claims $i + 1$ through 200 are true, but that we cannot know what number i really is—perhaps even that it is in principle impossible for anyone to know what i is, no matter their powers to collect and reason about the relevant evidence. However, this leaves plenty of room for arbitrariness. Whether we know it or not, if the vague denotation of CAT is determined by some particular 'choice' of i, then different choices will result in different vague denotations, and so different concepts, but many of these concepts will be just as useful, and just as otherwise distinguished, as CAT. Epistemicism does not affect the complexity of the semantic value of CAT when compared to considering CAT as a non-vague concept, and it was that complexity that allowed us to generate so many candidate denotations among which we could not usefully distinguish. So epistemicism does not resolve the arbitrariness of CAT.

A different theorist might claim that truth comes in degrees and that assertions involving vague terms can have one of potentially infinitely many truth values—if truth values are modeled using real numbers, 0 might represent false, 1 true, and every number between 0 and 1 some degree of truth between false and true.[15] When it comes to the 200 claims involving the vague term "cat" under issue, then, this kind of theorist would say that there is some pair of numbers, i and j, such that $1 < i < j < 200$, claims 1 through i are fully false, claims j through 200 are fully true, and from claim i to claim j the truth gradually increases.

This kind of response might seem to take the arbitrariness of CAT more seriously than does epistemicism, since this response does not posit a single precise boundary around the denotation of CAT. However, recognizing that CAT is vague and applying a degrees-of-truth theory to it still does not resolve CAT's arbitrariness. Instead of the single arbitrary choice for i we had with epistemicism, we are at best left with two arbitrary choices, i and j. There will be at least as many undistinguished choices for the i-j pair as there were undistinguished choices for the single boundary i. This kind of degrees-of-truth approach to predicates simply involves a more complex boundary than one expects to find in non-vague predicates, and with increased complexity typically comes increased arbitrariness.

Lastly, a supervaluationist[16] might claim that there are many ways we might acceptably precisify "cat"—many ways we might take the vague term "cat" and replace it with a non-vague but otherwise similar term. And when it comes to the series of 200 "cat"-involving claims, claim 1 is false under every acceptable precisification, claim 200 is true under every acceptable precisification, and then there is some continuous subseries somewhere in the middle of the larger series such that every claim in the subseries is true under some acceptable precisifications and false under some acceptable precisifications. But a cagey supervaluationist will claim that "acceptable precisification" is vague just as "cat" is, and so will deny that there is any particular pair of i and j precisely marking the boundaries of the continuous subseries of claims which are true under some, and false under other, acceptable

[15] Smith (2008), for example, claims that only theories which have truth coming in degrees can correctly handle vagueness.

[16] For a classic example of supervaluationism, see Fine (1975).

precisifications of "cat."

Since a supervaluationist can avoid this precise choice of i and j, unlike the other theorists discussed above—since a supervaluationist denies that any single precisification of a vague term is specially distinguished—supervaluationism, among the three theories of vagueness I am discussing here, might seem uniquely able to resolve the arbitrariness of CAT. However, even though a good supervaluationist (quite reasonably) seeks to avoid precision, she still leaves room for arbitrariness, because the arbitrariness of CAT is prior to any moves to precisify it, just as HEAP has arbitrariness even when considered as a vague concept.

To see how early the arbitrariness of CAT comes in, let us return to the 200 properties which we are considering as candidates for the denotation of CAT, c_1, c_2, ..., c_{199}, c_{200}, with c_1 being a poor candidate and c_{200} being a good candidate. We can call the dimension along which they vary their f-ness, so that for n from 1 to 199, c_{n+1} differs from c_n only in being slightly more f-ish. Now let us compare the concept CAT to a concept which is just like CAT but has the same denotation as EITHER CAT OR CAT-LIKE BUT VERY SLIGHTLY LESS F-ISH CREATURE. The two clearly have different denotations—the latter being very slightly more permissive—and this should be reflected in the supervaluationist's analysis, even if the supervaluationist need not identify a precise number of stops along the 200 degrees of f-ness by which the two denotations differ (since VERY SLIGHTLY, which fixes the difference along the dimension of f-ness by which the two vague denotations differ, is plausibly itself vague). But if we have chosen an appropriate f[17] , there will be no practical difference between the two (vague) denotations: each could serve as the denotation of CAT just as well as the other. The only thing distinguishing them is our assumption that one happens to be the true denotation of CAT, but this makes no practical difference. If instead of CAT we had a concept with the other denotation, we would get along just as well in our practices as pet owners, internet video viewers, and zoologists.

So CAT involves arbitrariness whether we are considering it as involving a choice of one from many candidate precise denotations or as involving a choice of one from many candidate vague denotations. Thus, whether we prefer a theory of vagueness which allows us to analyze away any semantic imprecision or which allows us to hold on to semantic imprecision all the way down (and up), we cannot simply resolve the arbitrariness of CAT by recognizing that it is a vague concept.

2.3 Where there is arbitrariness, publicity needs explanation

So the denotation of CAT is arbitrary, even if it is vague. This is relevant to publicity because given my assumptions, on individualistic theories of concepts the question of whether two

[17]The freedom we have in choosing f serves to further illustrate the arbitrariness of CAT: not only does precisifying CAT involve choosing a denotation with the right degree of f-ness, where there will be many candidate degrees, but given the complexity of CAT we should not be surprised if there are many other dimensions along which its denotation might vary, finely if not continuously.

people share the concept CAT depends on the question of whether the one possesses a concept with the denotation of CAT and the other also possesses a concept with the denotation of CAT. And if the denotation of CAT is arbitrary, we have no antecedent reason to expect either person to possess a concept with that denotation, rather than possessing a concept with a subtly different denotation.

I briefly made this point when discussing the case of which side of the road we drive on above, but a comparison to another case of obvious arbitrariness might help to make this point clear, especially if that case is more like the case of the denotation of CAT than the case of which side of the road we drive on.

So: the sound of our word for cats is arbitrary. English speakers could have settled on one of any number of sounds other than the one we actually use to talk about cats without compromising our ability to do so, and the same goes for other linguistic communities (except perhaps any using an onomatopoetic word for cats). In light of this, if we are told that person A is talking about cats and that person B is talking about cats, we have no reason to conclude that they are using the same word (assuming that we aren't given any other information and that words are individuated at least according to their sounds). Indeed, if we were to randomly and independently sample two people from everyone on Earth talking about cats, if anything we should expect them to be using different sounds.

Of course, there are some important differences between words' sounds and concepts' semantic values. One obvious difference is that sounds are publicly directly accessible: we can just hear what sounds other people use. This fact might give us some fairly straightforward explanations of how, in spite of arbitrariness, we manage to share words throughout large linguistic communities.[18] But even if these explanations are straightforward, they are necessary if we are ever to have well-founded expectations of coincidence of different people's sounds for cats. In light of the great variety of sounds perfectly suitable for use in talking about cats, in a case in which two people are talking about cats, absent any explanation for their using the same word (such as their both having grown up in the same linguistic community, hearing people use the same sound to talk about cats), we should not expect them to use the same word.

And just as we should not expect two people to share a word unless we have some particular explanation of their doing so, if that word has an arbitrary sound, we should not expect two people to share a concept unless we have some particular explanation of their doing so, if that concept has an arbitrary denotation. As I will discuss in the next section, giving such an explanation from within an individualistic theory will not be trivial.[19]

[18]It is worth noting here that although I think the public accessibility of words is relevant to their publicity, I cannot see any unproblematic simple accounts of the publicity of words. As I will discuss in a later chapter, I think individualistic accounts of word possession (on which the question of whether two people possess the same word in their vocabulary boils down to the more basic questions of whether the one person possesses a word with a certain sound and meaning and whether the other person possesses a word with the same sound and meaning) turn out to be just as problematic with respect to publicity as individualistic accounts of concept possession, and for similar reasons.

[19]Although he is discussing words, rather than concepts, and he does not attach the same importance to

Before I take up the question of whether individualistic theories of concepts leave room for explanations of concept sharing, though, it is perhaps worth pausing to again distinguish arbitrariness from indeterminacy. I am not claiming that in order to allow for publicity, individualistic theories of concepts need to (although they cannot) provide tools that allow us to explain how each thinker acquires the particular concepts she does—that individualistic theories of concepts need to (although they cannot) give determinate answers about which concepts thinkers possess. I do not need this claim for my argument, and I do not think it is true.

As far as I am concerned, on an individualistic theory of concepts, there may well be facts of the matter about what concepts thinkers acquire. If the theory is any good, the theory will surely pick out some features of a thinkers and use the features to somehow explain why the thinker acquires the concepts she does. If the theory picks out features which are subtle enough, it might be able to distinguish between the subtly different candidate concepts and give a verdict about exactly which particular concepts the thinker acquires, avoiding indeterminacy. But in the absence of some kind of special explanation of concept sharing, which might have to go beyond explanations of concept possession, when looking at multiple thinkers and asking which CAT candidate each thinker possesses, we will have no reason not to expect some kind of random distribution across the thinkers, given that we have no reason not to expect a similar distribution across the thinkers in the subtle features that, on the theory in question, make the difference between acquiring one candidate concept rather than another.[20]

The question, then, is: what explanations of concept sharing do individualistic theories of concepts offer, or even leave room for?

---

semantic value that I do, Noam Chomsky seems to recognize the importance of this kind of explanation to individualistic theories like his when he makes a poverty-of-the-stimulus-style argument for the innateness of universal "semantic properties" of some kind:

> The study of meaning has to face the fact that extremely limited exposure in highly ambiguous circumstances suffices for children to come to understand the meanings of words and other expressions with remarkable delicacy, far beyond anything that the most comprehensive dictionaries and grammars begin to convey, with refinements and intricacy that are barely beginning to be understood. For such reasons, empirical inquiry has sought to discover semantic properties that are innate and universal. (Chomsky, 2000, 185)

There is surely something to Chomsky's argument. However, in the present dialectical context the suggestion that there are innate, *universal* semantic properties that can somehow allow for an explanation of concept sharing from within an individualistic theory simply amounts to a more empirical analogue of the metaphysical claim of 'sparse' natural properties discussed above in this chapter. It is far from clear that individualistic theories of concepts ought to depend on such a controversial claim for their compatibility with publicity.

[20]It is perhaps worth returning here to the comparison to coordination problems that I made in an earlier footnote. In the case of a coordination problem where each player has a truly arbitrary choice from a common set of options, it can still be that each player chooses some option or other. But because the choice is truly arbitrary—more precisely, because we have no explanation of coordination—the only outcome we can expect is a random distribution of the options across the players.

2.4 Individualistic theories of concepts cannot explain how people share CAT

This is the premise of the argument which is most difficult to support in its general form, because individualistic theories of concepts admit of significant variation. In the next chapter, I will try to show that two of the most prominent families of individualistic theories of concepts—causal and inferential-role theories—are not compatible with explanations of concept sharing. For now, though, the most I can do is give some sense of the difficulty of the task before these theories, and then make some general remarks about the resources these theories have available to help them to accomplish this task.

Suppose we can somehow reduce the arbitrariness of CAT's denotation, *cathood*, to being an arbitrary choice between one hundred good candidate properties. This is surely very ambitious, given the complexity of *cathood*,[21] but it is still supposable. And further suppose that we are considering two individuals, A and B, that A definitely possesses the concept CAT, and that B possesses a single concept which serves as well as CAT—it might be CAT, or it might be a concept which is like CAT but which has one of the other ninety-nine candidate denotations. We are interested in the question of whether A and B share CAT.

Now, as I discussed in chapter 1 and briefly mentioned above, the defining feature of individualistic theories of concepts is that they have us take up the question of which concept it is that an individual possesses before taking up the question of with whom that individual shares that concept. So, on an individualistic theory of concepts which individuates concepts at least according to their semantic values, the question of whether A and B share CAT will come down to the more basic question of whether the concept B possesses denotes *cathood* (perhaps in addition to other more basic questions). If we want to explain how A and B do indeed share CAT, then, we'd need to explain how B manages to hit upon the 'right' denotation out of the hundred good candidate properties, keeping in mind that A and B probably never find themselves in situations which would allow distinguishing any of the good candidates from one another.

If all of the candidate denotations are really not distinguished from one another with respect to B, then we should expect which denotation B's concept has to be a simple matter of random chance. And if there is nothing more to being the concept CAT than denoting *cathood*, then we should expect that there is a one in one hundred chance that B possesses CAT, and so shares CAT with A. Even making some extremely optimistic assumptions, then, it seems that it is only with a fair bit of luck that A and B would share CAT. This is, of course, not to say that they could not share CAT, or that given the large number of people in B's situation that there would not be a few who, on some individualistic theory or other, come out luckily sharing CAT with A. But a few people sharing CAT by luck is not enough to make CAT public in the most relevant sense.

[21] Given the many dimensions along which properties close to *cathood* might vary, and the many degrees to which they might vary along these dimensions, we should expect the number of good candidate properties to be much larger than one hundred, if not infinite.

So it seems that any explanation of A and B sharing CAT other than by luck would have to involve a denial of the above assumption that none of the candidate denotations are distinguished from one another with respect to B. To deny this while admitting the arbitrariness of *cathood*, a theorist would seem to have to bring in the fact that A (or someone else around B) possesses CAT, which denotes *cathood*. But how is this fact relevant to B? One obvious way that it might be relevant is that there might be some reason to say that A and B share CAT. But on an individualistic theory of concepts, this fact depends on the more basic facts about which concepts A possesses and which concepts B possesses, so to say that B possesses CAT in part because B shares the concept with A would be to reverse the allowed order of explanation.

Another option—one which is not immediately ruled out by individualistic theories of concepts—would be to say that B possesses CAT in part because it is the concept that A possesses, and B's concept possession somehow depends on A's concept possession, which is somehow prior to B's concept possession. Explanations of this form are often brought in in individualistic theories of concepts, especially if B is in a position to defer to A in using the concept CAT or an associated natural-language term like "cat." Peacocke, for example, offers an explanation like this, and I will discuss it in the next chapter. However, it is worth saying now that if it is to make room for a publicity of concepts as widespread as we ordinarily think of it, deference has a great deal of work to do, and it is far from clear to me how deference can be made to do this work while holding on to individualism in a theory of concepts, especially given how much we use concepts and terms without deferring to others.

On an individualistic theory of concepts, then, the other place to look for the basis of an explanation for B possessing CAT instead of a very similar concept is among the facts that are explanatorily prior to A possessing CAT. There are presumably some features of A in virtue of which A possesses CAT instead of a merely similar concept, perhaps including A's interactions with cats or the inferences A makes when confronted with cats. If we are trying to explain how A and B share CAT, we might hope to find some way in which A's having these features is relevant to which concept B possesses—perhaps B's interactions with cats are relevant to which concept B possesses, and A is sometimes involved in B's interactions with cats, so that B's interactions sometimes involve A's interactions. Or we might to find some way in which B has the same concept-determining features as A—perhaps there are some parts of human nature that single out CAT from its close neighbors, and so CAT is not really as arbitrary with respect to A and B as it at first might seem. Fodor offers explanations along the lines of both of these, and I will discuss them in the next chapter. However, it is again worth saying now that I do not find these explanations plausible, at least if they are meant to secure a publicity of concepts which is as widespread as we ordinarily think of it.

Although it is not directly relevant to establishing this premise, or indeed to making this chapter's argument, it might be helpful to make a brief comparison to social theories of concepts here, since I ultimately hope to use this chapter's argument to motivate social theories by bringing out a serious problem with individualistic theories. The reader will recall that social theories take the question of which thinkers share a concept to be prior to

the question of which concept that concept is, a question that includes the question of what that concept's semantic value is.

On a social theory of concepts, then, the case in which A possesses the concept CAT and B might or might not share that concept with A is a bit strange, since whether B shares that concept with A might be one of the features of that concept that decides whether that concept is CAT or some other concept. But ignoring that strangeness, it should be fairly clear that if a social theorist wishes to explain how A and B share CAT, she has available to her a kind of strategy that the individualistic theorist does not: she can try to explain how A and B share the concept without first (and independently) having to explain how B possesses CAT.

I do not want to devote much space here to discussing exactly how the social theorist might explain how A and B share a concept, since I will take up that question at great length in a later chapter. But I do want to try to give some idea of how the social theorist's task might be feasible where the individualistic theorist's is not.

In order to say that A and B share the concept CAT, it will be enough if the social theorist picks out some relation between A and B that is somehow specific to A's concept CAT and to B's concept the identity of which is in question. For example, a social theorist might say that if A expresses some concept with the word "cat" and B takes all A's "cat"-involving statements, when B hears them, to be relevant to B's thoughts involving the concept in question, then A and B thus share a concept, whatever that concept is, so if A's concept is CAT then A and B share CAT. This is, of course, much too quick of an explanation, but it is a kind of explanation that is available to a social theorist and not to an individualistic theorist. An individualistic theorist wishing to give an explanation along these lines would have to explain how B's taking A's "cat"-involving statements to be relevant to B's thoughts involving the concept in question could distinguish CAT from the a subtly different concept that A would express with identical-sounding statements if A possessed that concept instead of CAT. This is because the individualistic theorist would have to explain how B possesses CAT (rather than a subtly different concept) without first explaining how A and B share a concept, because the individualistic theorist takes B possessing CAT to be more basic than A and B sharing a concept.

Individualistic theories of concepts, then, are significantly hindered when it comes to giving explanations of concept sharing, at least when it comes to complex concepts with arbitrary denotations like CAT. Given just how arbitrary CAT's denotation is, this might seem like an insurmountable problem for individualistic theories. In case it does not, in the next chapter I will take up Fodor's and Peacocke's theories of concepts as representative examples of individualistic theories in the hopes of establishing beyond any doubt that individualistic theories cannot overcome this arbitrariness. But before I move on to that discussion, I will take up the final premise in my argument and try to give a clear sense of just how widespread this arbitrariness is among our concepts.

2.5 CAT is not special in being like this

A significant proportion of the above argument did focus on features of CAT which are particular to it, or at least to many concepts of biological species (or clades). And I have already suggested that there are some other concepts, like WATER (at least if WATER has the specific and simple denotation *being composed of* H_2O), to which this argument cannot be applied. So how widely can the argument be applied?

It seems to me that practically any sufficiently complex general concept falls within the scope of this argument. The feature of CAT that was necessary for the argument was just that it had a denotation such that there were many other denotations differing from it only by subtle variations and not differing significantly enough to matter for the uses to which we put CAT. This is all that is required.

CAT, like many species concepts, offers many dimensions along which a denotation might vary subtly, but even simpler concepts can be complex enough to have this kind of arbitrariness. For example, suppose that color concepts' denotations differ only along two dimensions: the lower and upper bounds of the portion of the light spectrum with which that color is (somehow) identified. Even with this extreme oversimplification, color concepts admit of enough subtle variation to be relevantly arbitrary. For comparing someone possessing the concept RED with someone possessing a concept which is just like RED except that its boundaries are shifted a single nanometer up or down the spectrum, we would not likely see any differences in behavior except in fairly uncommon and usually unimportant circumstances—and indeed, we should expect to see differences like these anyway in people who we ordinarily think of as sharing RED, since disagreements about observed colors are, even if uncommon, hardly unfamiliar. Even focusing only on spectrum boundaries, RED and other color concepts feature enough arbitrariness to fall within the scope of the argument.

The same will fairly clearly go for vague natural-kind concepts like TREE and PLANET (and WATER if it has a denotation more like *mostly being composed of* H_2O), complex functional concepts like HOUSE and STUDENT, abstracted social or normative concepts like POPULAR and JUST, and a great many other general concepts as well. It will also go for many relational concepts, such as KNOWS, EATS, and INSIDE, and many concepts for events and actions, such as ELECTION and SLEEPS. Indeed, most vague concepts should serve, since vagueness is usually the practical result of recognizing arbitrariness.[22]

[22]Singular concepts of complex objects can plausibly also involve a kind of arbitrariness which brings them within the scope of this argument. If singular concepts denote objects, and many of these objects (including particular people, buildings, and geographical features) are individuated according to their microphysical constituents or space-time boundaries, then in many cases subtle variations of these constituents or boundaries will make for subtly different but practically indistinguished singular concepts. Just as with many general concepts, we might wonder why two thinkers would both happen upon the concept of a particular object when there are so many very similar objects each might have thought about instead.

This is, of course, an important issue in its own right, closely connected to the "Problem of the Many" discussed by Unger (1980) and in the literature responding to that article. However, this case is if anything more metaphysically ramified than the case of general concepts, and I do not have the space here to give it the treatment it deserves.

The only concepts which fall outside of the scope of the argument, then, are concepts which are not arbitrary in any way, or which only involve arbitrary selection among possibilities which are either few in number or not so similar as to be practically indistinguishable. ELECTRON is perhaps an obvious example of such a concept, as are similarly simple natural-kind concepts and many mathematical and logical concepts, because their denotations do not fall along dimensions that admit of continuous or nearly-continuous variation. Metaphysical concepts like TIME and SPACE might be somewhat less obvious examples, since even if they are more complex, it is not clear that we could just as easily think using alternatives.

This is not to say, though, that there is no reason to worry about the publicity of ELECTRON or TIME. If concepts are individuated according to more than just their semantic values, these other individuating features might make trouble for the publicity of even these concepts. For example, if concepts are individuated according to their total role in inference in an individual's mind, or if they are individuated according to something like a mode of presentation, there might be reason to doubt that two people share the concept ELECTRON even if what each is thinking about does turn out to amount to *electronhood*. I will just not be taking up such worries in this work.

2.6 Therefore, individualistic theories of concepts have a problem allowing for publicity

So, to re-summarize my argument:

The concept CAT has an arbitrary denotation. This is because CAT has a fairly complex denotation, so that there are many denotations which are like the denotation of CAT but only very subtly different. It is also because the standards motivating us when we acquire CAT, rather than a concept with a subtly different denotation, are not fine-grained enough to distinguish among many of these denotations.

The arbitrariness of the denotation of CAT stands even if we hold that it features some kind of vagueness, because vagueness does not eliminate arbitrariness—it is merely a feature of language (and perhaps thought) that allows us to continue to use terms while recognizing their arbitrariness. Similarly, this arbitrariness stands even if we hold that one potentially important factor in assigning a denotation to CAT is the naturalness of the property denoted, because naturalness does not distinguish any one of the many candidate denotations for CAT—on any plausible account of naturalness, there will be many denotations that differ only subtly from the denotation of CAT and are just as natural.

Where there is arbitrariness, we should not expect widespread sharing unless we have an explanation for it. This is because in cases of true, unqualified arbitrariness, no option is distinguished from all others. If no option stands out, the only way many agents

will end up selecting the same option is by chance—or because the ratio of options to agents is so low that it is inevitable. So if we are to have a reasonable expectation that many thinkers will acquire the same concept, it must be because we can see that that concept does not involve true, unqualified arbitrariness—it must be because we have an explanation of widespread sharing that somehow singles out that concept. In the case of a concept like CAT that involves some significant arbitrariness, then, we would need an explanation that shows how the concept falls short of true, unqualified arbitrariness.

Individualistic theories of concepts cannot explain how many people possess CAT. This is because they are limited in the features of CAT they can call on in giving explanations of a given thinker's possessing CAT. In particular, these theories do not allow explanations of a thinker's possessing CAT that depend on the claim that that thinker shares the concept in question with another thinker, because the distinguishing feature of individualistic theories of concepts is that they take facts about concept sharing to depend entirely on facts about which concepts individual thinkers possess. The importance of this limitation of individualistic theories will become more clear in the next chapter.

This argument applies to other complex general concepts as well as CAT. This is because the only feature of CAT that this argument depends on is its arbitrariness, and other complex general concepts are arbitrary in the same way. All that is required is that there be many denotations that differ only subtly from the concept's actual denotation and that these denotations are not practically distinguished from one another. There are relatively few concepts that do not seem to exhibit this arbitrariness—among them are mathematical and logical concepts, as well as concepts of simple natural kinds like ELECTRON.

Therefore, on individualistic theories of concepts we should not expect many people to share complex general concepts. This, then, is the problem of publicity: publicity of concepts—of concepts like CAT and RED, as well as of simpler concepts—is an important desideratum for a theory of concepts, and individualistic theories of concepts cannot allow for publicity, even if that publicity only involves many people actually sharing the concepts in question. This gives us a strong reason to adopt a different kind of theory, viz. a social theory of concepts. I will take up the topic of social theories of concepts later in this work, developing a sketch of such a theory and showing how it avoids the problem of publicity.

I hope that this argument is already persuasive. However, I expect that I have not yet resolved all doubts about whether individualistic theories have the resources to explain how a concept like CAT is widely shared. In particular, I have not yet discussed in detail any well-worked-out individualistic theories, and I have not yet discussed at all any of the moves that individualistic theorists have actually made that might seem to help them account for publicity in the face of a challenge like mine. So, in the next chapter, I will take up in turn Fodor's and Peacocke's theories of concepts, making clear how each theory in itself has

a problem accounting for publicity and also making clear how each theorist's attempts to account for publicity cannot succeed, at least if they are to maintain their individualism.

Chapter 3

Causal and Inferential-Role Theories

This chapter continues the argument made in the last chapter. In that chapter, I argued that individualistic theories of concepts are not able to allow for the publicity of concepts, but I did not discuss any particular detailed individualistic theories. Instead, I made the argument more generally. In this chapter, I will apply the argument to two particular individualistic theories of concepts, showing how they—and other theories like them—are incompatible with the publicity of concepts, even in spite of their authors' efforts to allow for publicity.

Before I get to the arguments, though, I should remind the reader about some of the upshots of the previous chapters' discussions.

First: When I talk about individualistic theories of concepts, I mean to talk about theories that take who shares a concept to be a less basic feature of that concept than which concept that concept is, including what that concept's semantic value is. So, for example, if it is stipulated that the thinker A possesses the concept CAT, which denotes *cathood*, and the question of whether the thinker B shares CAT with A is raised, then on an individualistic theory the answer to this question would depend on the answer to the more basic question of whether B possesses the concept CAT, which would depend in part on whether B possesses a concept that denotes *cathood*.

Second: When I talk about the publicity of concepts, for the purposes of this chapter I just mean to talk about many people actually sharing many concepts as we ordinarily take them to. This sharing should not be limited to simple concepts like AND and TWO; complex but ordinary concepts like CAT and RED should be widely shared as well.

Third: The complaint I made about individualistic theories in the last chapter is that they cannot overcome the arbitrariness of complex general concepts like CAT by explaining how many thinkers could share CAT—or otherwise share a concept in the neighborhood of CAT—rather than each possessing one of the many concepts with similar but subtly different denotations. I will not go back over all of the last chapter's argument here, but the thrust of the argument was that given the arbitrariness of CAT, without some special explanation—which individualistic theories cannot offer—we should not expect that many thinkers will share CAT, and we should instead expect that different thinkers will possess different concepts that are only similar to CAT, so that no single concept in the neighborhood

of CAT will be widely shared. In other words, given the arbitrariness of CAT, publicity needs to be accounted for if we are to expect it.

The main aim of this chapter is to drive home the claim made in the last chapter that individualistic theories cannot account for the publicity of concepts like CAT. Taken together with the conclusions of the last chapter, this implies that on individualistic theories, we should expect that concepts will not be public. In the next chapter, I will discuss at length why publicity—many people having many of the same concepts as many other people—really is an important desideratum on a theory of concepts, which has up until now only been assumed. These three chapters together, then, will establish that individualistic theories of concepts come with a serious cost. It will only be in the final chapter, though, that I will sketch the development of a social theory of concepts so that the costs and benefits of individualistic and social theories can be thoroughly compared—for now I am only trying to establish that individualistic theories of concepts cannot account for publicity.

So in this chapter, I will take up the question of whether individualistic theories of concepts really can offer explanations of widespread sharing of concepts like CAT. I do not have the space to take up every live individualistic theory, so I will focus my efforts on just two theories. I have selected these theories for several reasons.

One reason is that I think they are both well-worked-out examples of theories from their respective families of theories of concepts. Each theory is relatively clear, well-motivated, and detailed. As a result, there will be relatively few points in my discussion at which I will be forced to guess what response a proponent of these theories would give to my complaint.

Another reason is that these theories are from two fairly different and fairly popular families of theories of concepts: causal and inferential-role theories. There are, of course, many live theories that fall outside these families, and I would like to be able to discuss them all. But the scope of my project is limited, so I will focus on these two families. Being fairly different from one another, if my argument can apply to both of these families then it should stand a good chance of applying to others—indeed, between the general argument from the last chapter and the two theory-specific arguments from this chapter, it should be fairly clear how to extend my argument to apply to any other individualistic theory of concepts. Additionally, being fairly popular, if my argument can apply to both of these families then it will be important even if it cannot be applied to others as successfully.

Lastly, these theories are particularly interesting because the theories' authors both attempt to give explanations of publicity that are compatible with their theories. That is, they both attempt to respond to something like the complaint I am making. I think these attempts help to illustrate the limitations of individualistic theories in general, especially because they are better-developed than some attempts other individualistic theorists make.

So in the next section, I will discuss Fodor's theory of concepts as an example of a causal theory of concepts. And in the section after that, I will discuss Peacocke's theory as an example of a conceptual- or an inferential-role theory.

3.1 Causal theories of concepts

Individualistic theories of concepts offer possession conditions for concepts—for each concept, the conditions a thinker meets insofar as she possesses the concept—and identify two thinkers' sharing a concept with those thinkers' both meeting the possession conditions for that concept. One of the main differentiators of individualistic theories of concepts is the form, on each theory, the possession conditions of concepts tend to take.

In the family of theories I am discussing in this section, 'causal' individualistic theories, possession conditions involve a strong causal component: in order to possess a certain concept, a thinker (or perhaps some part of the thinker) must be involved in a certain kind of causal relation, perhaps a kind of causal law, typically to some property or object. For example, in order to possess the concept CAT a thinker might have to be causally related to *cathood* in a certain way, because being involved this causal relation determines the denotation of the concept in question, and to be the concept CAT a concept must at least denote *cathood*.

The distinguishing feature of causal theories of concepts as a family is the claim that a concept's semantic value is determined by the concept's involvement in a certain special causal relation, an instance of which is the relation *cathood* bears to CAT—typically, the concept's semantic value will be identified with the property or object to which the concept is thus related. It is worth noting that this claim does not imply individualism about concepts: one could hold that a concept's semantic value is determined by its causal relations but deny that that semantic value is a more basic feature of that concept than who shares that concept—especially if one held that that concept's semantic value is determined by the causal relations in which it is involved as it is possessed by all those thinkers who possess it. But I am not interested in a theory like this in this chapter, so for the rest of this section when I say "causal theory" I will generally be talking only about individualistic causal theories.

3.1.1 Fodor has a causal theory of concepts

Jerry Fodor's theory of concepts belongs in this family, because he individuates concepts[1] in part according to their semantic values, and he determines the semantic values of concepts according to some of the causal relations in which they are involved.

More specifically, Fodor assigns semantic values to concepts based on the information they carry, for which Fodor offers a causal analysis. For example, you can use your concept CAT to think about cats, on his view, because your using the concept carries information about cats, and your using the concept carries this information because your using it tends to be caused by cats. This is Fodor's flavor of informational semantics at its most basic.

[1]Actually, Fodor treats concepts as mental particulars, each a token of some concept type. So strictly speaking, his theory should type concepts, rather than individuating them. But Fodor himself uses the language of individuation, and it is easy enough to translate a theory of (concrete) concept typing into a theory of (abstract) concept individuating, so I will do so here for the sake of consistency with the rest of my discussion.

Fodor neatly puts "the Ur-version" of the theory in "Information and Representation": "As carry information about Bs iff the generalization "Bs cause As" is true and counterfactual-supporting" (Fodor, 1999, 514). Fodor also sometimes gives this basic part of the view by saying that the generalization must be a law[2], along the lines of "fire causes smoke" or "water flows downhill," or that the concept (or sometimes the thinker) is "locked to" the relevant property[3]. There is, of course, much more that needs to go into an information-based semantic theory if it has any hope of capturing the meanings of concepts or thoughts, but I will only bring in further details of Fodor's theory as they become relevant to my argument. What is important for now is that Fodor's theory is founded on the idea that concepts pick out the things which cause their use.[4]

3.1.2 Fodor's theory cannot overcome the arbitrariness of CAT

When it comes to the semantic value of the concept CAT, it is fairly clear that this basic version of Fodor's theory will say something along these lines: if an individual possesses a concept C, then C has the same semantic value as CAT just in case it is true and counterfactual-supporting that cats cause uses of C—or alternatively, just in case cats (as a group) and C are related by the right kind of law, or just in case C is locked to *cathood*. For the purposes of relating Fodor's theory to my argument in this chapter, I think it will be easiest to put things in terms of a concept or thinker 'locking to' a property, so I will generally stick to that terminology going forward.

In the context of Fodor's theory, then, the arbitrariness of CAT discussed in the previous chapter amounts to there being many properties very similar to *cathood* which are just as available for locking to as *cathood* is, with little or no practical difference among them all. The question of whether this arbitrariness can be overcome amounts to the question of whether there is any reason, consistent with the theory, to expect that different thinkers will somehow lock to the same property out of *cathood* and all its close neighbors, either because *cathood* really is special after all or because the thinkers will somehow coordinate their locking.

3.1.2.1 Problem cases: getting locked to *cathood*

Now, Fodor has certain commitments which place strong constraints on the kind of stories he can tell about how it is that we have the concepts we have, particularly when it comes to concepts like CAT which seem unlikely to be innate. As a consequence, over the years he has

[2]e.g. Fodor, 1998, 12

[3]e.g. Fodor, 1998, 125

[4]Of course, not *all* uses of a concept should determine its semantic value. For instance, if a dog disguised as a cat causes me to use my concept CAT, or if a dog which looks nothing like a cat causes me to use CAT in the thought THAT IS NOT A CAT, CAT should not thereby come to refer to the dog. Fodor has things to say about these kinds of cases, and some of these things will be discussed later, but they will not help him with the problems I am presenting here.

told some fairly controversial stories about how it is that we have concepts like CAT, stories which include explanations of how it is that we are locked to properties like *cathood*. However, the constraints within which Fodor is working and the development of these explanations are outside the scope of this work, since they need not be shared by everyone with a theory of concepts which determines semantic values according to causal relations in more or less the way Fodor's theory does. So the question I will take up in this subsection is: what kinds of stories can a causal theorist (of this kind) tell about getting locked to *cathood*, and do any of these stories allow the arbitrariness of CAT to be overcome?

Fodor on *doorknobhood* In his book *Concepts*, when discussing concepts which are not plausibly innate, Fodor tells a very short story about acquiring the concept DOORKNOB by getting locked to the property *doorknobhood*:

> ... *doorknobhood* is the property that one gets locked to when experience with typical doorknobs causes the locking and does so *in virtue of the properties they have qua typical doorknobs*. We have the kinds of minds that often acquire *the concept X from experiences whose intentional objects are properties belonging to the X-stereotype*. (Fodor, 1998, 137-8)

Given that Fodor thinks a great many of us share DOORKNOB, the basic idea seems to be that there is a single property, *doorknobhood*, such that almost any human encountering a few salient-enough typical doorknobs will get locked to that property. And we lock to that property, instead of some other property, because that property just is the property that almost any human encountering a few salient-enough typical doorknobs will get locked to.

So if Fodor is right about *doorknobhood*, and a similar story about *cathood* is true, then it turns out that *cathood* is not arbitrary after all, since it turns out that there is just the one property that humans get locked to upon encountering a few salient objects typical of the relevant kind. So even though a human encountering a few animals with the property *cathood* will thereby also be encountering a few animals with the properties $cathood_1$, $cathood_2$, and many other properties with extensions very close to *cathood*'s, that human will get locked to *cathood* because it just is the property that humans get locked to in those circumstances.

But, of course, one of the questions at issue is precisely the question of whether or not there is a single property to which humans get locked in these circumstances. The large number of practically indistinguishable (or at least undistinguished) properties in the neighborhood should cast doubt on the claim that there is such a single property, and Fodor's idea that that property might be identified as the property to which humans get locked in these circumstances rests on this claim but does nothing to help us overcome our doubts. Fodor does nothing with his story about *doorknobhood* to explain how so many humans manage to get locked to the same property out of all of the properties they might get locked to; he does nothing to overcome the arbitrariness of CAT.

Laurence and Margolis on kind syndromes It might be helpful, then, to see how other theorists working in a Fodorian framework might tell the story of getting locked to *cathood*, especially if that story goes in to more detail concerning the mechanism of the locking than Fodor's own story does. Luckily, Stephen Laurence and Eric Margolis tell just such a story in their article "Radical concept nativism." To tell their story, they first introduce the notion of a 'sustaining mechanism':

> A sustaining mechanism is a mechanism in virtue of which a concept stands in the mind-world relation that a causal theory of content, like Fodor's, takes to be constitutive of content. (Laurence and Margolis, 2002, 37)

Laurence and Margolis go on to describe a type of sustaining mechanism for natural kind concepts (like, say, CAT), involving a 'kind syndrome':

> A kind syndrome is a collection of properties that is highly indicative of a kind yet is accessible in perceptual encounters. This may include things like the typical shape, motions, markings, sounds, colors, etc., associated with a kind. ... This information, together with [a more general disposition to treat instances as members of the category only if they have the same essential property as paradigmatic exemplars of the syndrome], establishes an inferential mechanism that can explain why an agent tokens a given concept under the conditions which ... are constitutive of conceptual content. (Laurence and Margolis, 2002, 38)

So if the mechanism which sustains the link between CAT and *cathood*—the mechanism in virtue of which CAT denotes *cathood*—is supposed to be of this type, then we get a fairly plausible story of how a thinker who possesses CAT is locked to *cathood*.[5]

The story presumably goes something like this, if we call the thinker who possesses CAT 'A': A is locked to *cathood* in part because A is already capable of tracking the outward symptoms typical of *cathood* in perception. These outward symptoms include the shapes, sounds, and behaviors typical of cats. If A is a fairly sophisticated or appropriately situated thinker, the symptoms A tracks might also include a propensity to produce a certain allergic reaction in A, being called "cat" by someone speaking English, or other less immediately obvious properties. A then has two operative dispositions: first, to use (or 'token') CAT when encountering entities[6] with many symptoms in this collection, and second (capable of overriding the first in cases of known cat facsimiles), to use CAT when and only when

[5]I do not mean to suggest, of course, that this is the only plausible story of how thinkers lock to *cathood* or that Fodor is committed to this being how thinkers lock to *cathood*. Anyone with Fodor's theory clearly must have room for other locking mechanisms for concepts—even other locking mechanisms for natural kind concepts—because there are natural kinds like *electronhood* that do not plausibly have kind syndromes, and it is possible that whatever story Fodor would like to tell about locking to *electronhood* could also be told about locking to *cathood*. But this story offered by Laurence and Margolis does strike me as working particularly well for CAT and *cathood*, even if it does not clearly work for certain other concepts as well.

[6]The talk of using a concept when encountering entities with the relevant property is, of course, too simple here, unless "encounter" is construed extremely broadly, but it is not the important part of the story.

encountering something with the essence typical of entities with many symptoms in the collection. These two dispositions, together with the tracking of the right collection of symptoms, should ensure that A uses CAT only when encountering entities with *cathood*—that is, that A is locked to *cathood*.

The question now relevant, of course, is whether these dispositions really do ensure that A is locked to *cathood*. Of course, they might fail to ensure this if A's environment is not typical of actual human environments with respect to cats—say, if A is in an environment that features a great number of very convincing cat facsimiles. But cases like this are not relevant. The relevant cases are cases in which A is typical for a human with respect to cats, since what is at issue is whether a story like the one Laurence and Margolis offer can overcome the arbitrariness of CAT for typical—including typically situated—human thinkers.

So how does the arbitrariness of CAT, as discussed in the previous chapter, show up in Laurence and Margolis's story if A is a typical human thinker? How might A, instead of getting locked to *cathood*, get locked to a practically indistinguishable but distinct neighbor property like $cathood_1$ or $cathood_2$? It seems fairly clear that A could get locked to one of these neighbor properties if it exhibits with a subtly different kind syndrome and A tracks the outward symptoms involved in that kind syndrome instead of the symptoms involved in the kind syndrome of *cathood*.[7]

For example, if a neighbor property has all the same outward symptoms as *cathood*, except that the shape symptom includes slightly-too-doggy shapes, and A tracks these symptoms, then A will be locked to the neighbor property instead of *cathood*. A might track these slightly broader symptoms for any number of reasons, such as A's first cat being slightly more dog-shaped than average, or perhaps some subtlety of A's perceptual apparatus making A slightly less discriminating than average when it comes to shapes. There need even be no difference between A's actual behavior and the actual behavior of a thinker tracking exactly the symptoms of *cathood*—the kind of causal relation that Fodor and others take to be relevant here is at least partly counterfactual, so a difference between the symptoms that A tracks and the exact symptoms of *cathood* could be relevant here even if the slightly-too-doggy creatures that A would call cats are only merely possible.

More generally, if a neighbor property has outward symptoms which are subtly different from those of *cathood*, and A tracks these outward symptoms, then A will be locked to that neighbor property instead of *cathood*. And given the great potential for subtle variations in the outward symptoms of *cathood* which correspond to subtle variations in the essential property which produces the symptoms, there seems little reason to expect A to track precisely the outward symptoms of *cathood* instead of the outward symptoms of one of its many neighbors.

[7]It is also worth considering the possibility that there is no fact of the matter about which property A is locked to—that considered under certain conditions A seems to be locked to one property but that considered under other conditions A seems to be locked to another property, say. This leads back to the discussion of indeterminacy above, and as I pointed out in that discussion, while an appreciation of indeterminacy complicates the argument I am making, it does not clearly affect the outcome.

Causal theories like Fodor's, then, do not have the means to overcome the arbitrariness of CAT, at least when we consider their most basic versions. Of course, there is more to Fodor's theory than I have yet presented, and we might hope to find some means to overcome this arbitrariness in additions to the theory's foundations or in clever tricks that can be performed with the theory. I will discuss both of these hopes below. First, though, I want to show how the problems with publicity for Fodor's theory are even worse than they might already seem.

3.1.2.2 Problem cases: 'false' beliefs

I argued in the previous subsubsection that the most basic version of Fodor's theory cannot overcome the subtleties of the arbitrariness of CAT, because the subtle variations across thinkers will show up in different thinkers getting locked to subtly different properties. As it turns out, though, Fodor's theory cannot even handle much less subtle differences between thinkers.

First, I will cast the case in which two thinkers are locked to subtly different properties in a certain light, to allow comparison to a case in which two thinkers are locked to less subtly different properties. Suppose we are told about two thinkers, A and B, where A is locked to *cathood* and B is locked to some very subtly different neighbor property, but both speak English and use the word "cat" more or less normally. Intuitively, if we are asked to ascribe thoughts to A and B, then we can ignore this subtle difference. We will treat both A and B as having thoughts about cats, and unless we are told about any other differences between the two, we will treat them as having more or less the same thoughts. Because cases like this are so common, this is a potential problem for Fodor, because Fodor's theory would not treat the two as both having thoughts about cats.

Now, in this chapter I do not want to take up the question of whether we are correct in ascribing thoughts about cats to both A and B. That question is closely tied to the question of whether publicity is an important desideratum for a theory of concepts, and it is very important to my project, but it is not the subject of this chapter. Instead, I simply want to point out that we would intuitively describe the above case of A and B as a case in which two people have thoughts about cats which are more or less the same thoughts, and I want to contrast that case with another case which is also potentially problematic for Fodor. This other case is again again one in which, intuitively, both A and B have thoughts about cats. The difference is that, intuitively, A's beliefs about cats are true but some of B's beliefs about cats are false, and not subtly false either. The interesting thing about this case is that the feature of B that distinguishes B from A—the feature we would intuitively call B's false belief—is something which on Fodor's theory prevents B from having any thoughts about cats at all, whether true or false. This difference between the way we are naturally inclined to describe the case and the way Fodor's theory would have us describe the case makes neutral presentation somewhat tricky, but I hope the general idea will be fairly clear.

Imagine the following scenarios:

S1 A and B encounter a naturally furry cat. A says, "That's a cat there." B says, "Yes,

that's definitely a cat."

S2 A and B encounter a naturally furless cat. A says, "That's a cat there." B says, "No, it isn't. It doesn't have fur, and all cats have fur. I don't know what it is, but it definitely isn't a cat."

Further imagine that S1 and S2 are typical of A and B's encounters with cats: A tends to call all cats cats, but B only tends to call naturally furry cats cats. Apart from this anomaly, B's use of the word "cat" is normal, as is A's.

Now, let's assume that A and B's thoughts and concepts are working something like their utterances and words—in particular, that B has some concept MAYBECAT which B uses when making the judgment THAT'S A MAYBECAT THERE when encountering naturally furry cats but not when encountering naturally furless cats, and that A has the concept CAT which A uses when making the judgment THAT'S A CAT THERE when encountering cats of all kinds. As noted above, intuitively, B's abnormal stance towards naturally furless cats doesn't stop B from agreeing with A in B's assertion in S1, nor does it stop B from agreeing with A in B's judgment in S1, because B is still capable of talking and thinking about cats—intuitively, B's concept MAYBECAT is the same concept as CAT, and B just has the false belief that all cats have fur, or maybe even that being naturally furry is necessary for being a cat. The same goes for the apparent disagreements in S2.[8]

The question, of course, is whether MAYBECAT really is the same concept as CAT, as it would seem to have to be if A and B are to share thoughts. Unfortunately, if we use a causal theory like Fodor's to determine the denotation of MAYBECAT, the verdict is that MAYBECAT and CAT are different. This is because, at best, MAYBECAT is locked to the property *being a naturally furry cat*, rather than to *cathood*. After all, it isn't a law that cats cause B to use MAYBECAT; it's only a law that naturally furry cats cause B to use MAYBECAT.[9,10]

[8] As I will discuss below, our intuitions might change if we further imagine that B will continue to claim "All cats have fur" no matter what evidence or persuasion B is offered. But for the purposes of this example, we can assume that B, like A, is just as susceptible to correction as any ordinary person is.

[9] Of course, in S2 a naturally furless cat does cause B to use MAYBECAT in the thought THAT'S NOT A MAYBECAT THERE, but if Fodor's theory allowed this kind of use to determine the denotation of MAYBECAT in the same way that use in the thought THAT'S A MAYBECAT THERE did, the theory would have far worse problems than conflicting with publicity, since dogs and many other non-cats would, in many normal situations, cause B to judge THAT'S NOT A MAYBECAT THERE.

[10] Also, the idea (discussed in the last chapter) that more natural properties are, all things being equal, more eligible to be semantic values of concepts might seem relevant here, since *being a naturally furry cat* is likely less natural on many theories of naturalness than *cathood*. However, this idea should only be applied when there are two properties which are otherwise equally good candidates to be the semantic value of the term or concept in question, which is not the case here: *being a naturally furry cat* is the property that is nomologically linked to MAYBECAT, and *cathood* is not. Furthermore, if Fodor's theory were amended to give naturalness a higher priority in determing semantic values than nomological relations, *cathood* would be a better candidate than *being a naturally furry cat*, but *electronhood* would be a better candidate still, which is clearly not the right answer.

And although this CAT/MAYBECAT case is somewhat contrived, there is reason to think that there will be many real-world cases with the same structure: cases in which different thinkers are prevented from getting locked to the same property because one has what we would intuitively call a false belief that certain entities do not have that property. For example, if a substantial minority of US citizens refuse to call Barack Obama a Christian in spite of the evidence presented to them, instead insisting that he is secretly a Muslim, then they probably have a concept such that most of the same things cause them to use it as cause the rest of the citizenry to use the concept CHRISTIAN but Barack Obama doesn't, so on Fodor's theory it will—at best—come out that they have a concept which denotes all Christians other than Barack Obama, whereas the rest of the citizenry will have a concept which denotes all Christians. And although most of these publicity-preventing beliefs will be held only by a minority of thinkers, so many of us have so many idiosyncratic but unimportant beliefs with the right structure that we will, on Fodor's theory, share many fewer concepts with other thinkers than we would ordinarily take ourselves to, even ignoring the subtle differences among thinkers that are the focus of my more general argument in this chapter and the last.

So thanks to the prevalence both of false beliefs of a certain kind—beliefs like the belief that all cats have fur or that Barack Obama is not a Christian—and of subtle differences among thinkers' dispositions to use concepts, the most basic version of Fodor's theory has a problem accounting for publicity. But what about the full theory in all its sophisticated glory?

3.1.3 Asymmetric dependence does not help Fodor's theory account for publicity

One of the most important parts of Fodor's semantic theory—and a part that wasn't in the presentation above—is the way he allows for mistaken applications of a concept, which involves a phenomenon he calls "asymmetric dependence."

Imagine that I'm trying to count the cows in a barn: I see a cow and say "That's one cow," then see another cow and say "That's a second cow," then hear a somewhat cow-ish noise from the shadows at the back of the barn and, taking the noisemaker for a cow, say "That's a third cow." Unbeknownst to me, the noisemaker is actually a cat, but this cat has caused me to say "cow," so it looks like if Fodor were to use his semantic theory to assign a meaning to "cow"[11] the word would mean *cow or cat*, since both cows and cats cause me to say it. So, on Fodor's theory as described so far, it's going to be very hard for me to misapply a word or concept, since anything that causes me to use it would fall under its meaning. This would be a very serious problem, but Fodor has an answer: he can distinguish

[11] I am really interested in Fodor's semantic theory of concepts, not his semantic theory of public-language words, but his clearest presentation of asymmetric dependence is put in terms of words like "cow," rather than in terms of concepts like COW, so for the sake of consistency I am putting this discussion in those terms. The application to concepts should be clear enough, though.

between normal and mistaken uses (or, as Fodor would put it, between normal and mistaken 'tokens'):

> Cows cause "cow" tokens, and (let's suppose) cats cause "cow" tokens. But "cow" means *cow* and not *cat* or *cow or cat* because *there being cat-caused "cow" tokens depends on there being cow-caused "cow" tokens, but not the other way around.* "Cow" means *cow* because ... noncow-caused "cow" tokens are *asymmetrically dependent upon* cow-caused "cow" tokens. "Cow" means *cow* because *but that "cow" tokens carry information about cows, they wouldn't carry information about anything.* (Fodor, 1990, 91)

The idea seems to be that it's only because the cat takes advantage of my cow-detecting abilities—because it makes a somewhat cow-ish noise, or casts a somewhat cow-ish shadow, or something—that it causes me to say "cow," but the cows aren't taking advantage of my cat-detecting abilities when they cause me to say "cow," so there is an asymmetry between cats and cows which allows the theory to rule out cats as falling under the extension of "cow." Fodor sometimes (reluctantly) spells out this asymmetry in terms of counterfactuals (Fodor, 1990, 95): if cats didn't cause "cow" uses then cows still would, but if cows didn't cause "cow" uses then cats wouldn't either.

Since asymmetric dependence is designed to keep mistaken uses from determining the meaning of a word or concept, and at least some of the problem cases described above involve intuitively mistaken beliefs which might play a role in determining the meaning of a concept, a proponent of Fodor's theory might reasonably wonder whether asymmetric dependence can handle those cases. Unfortunately, it can't. Briefly, the reason that it can't help is that it is used in the theory to rule out mistaken uses from determining the meaning—to make sure the extension doesn't include these extra objects—but the problem cases presented above, at least those involving mistaken beliefs, are ones in which a refusal to use (or to use in the right way) plays the interesting role in determining the meaning—in which the problem is that the extension doesn't include the relevant objects.

In a bit more detail: in the CAT/MAYBECAT case presented above, if MAYBECAT is to denote *cathood* instead of *being a naturally furry cat*, asymmetric dependence could help if naturally furless cats causing B to use MAYBECAT asymmetrically depends on cats causing B to use MAYBECAT—that is, if the furless-cat-MAYBECAT link is distinguished because it depends on the cat-MAYBECAT link while the cat-MAYBECAT link does not depend on it, just as the cat-"cow" link is distinguished because it depends on the cow-"cow" link while the cow-"cow" link does not depend on it.

The problem, though, is that there isn't a law that cats cause B to use MAYBECAT, so naturally furry cats causing B to use MAYBECAT can't depend on this law. It may be true, in some sense, that cats cause B to use MAYBECAT: after all, most cats are naturally furry, so most cats do cause B to use MAYBECAT (when B thinks about them). But it's only true in this sense that cats cause B to use MAYBECAT because naturally furry cats do: if naturally furry cats didn't cause B to use MAYBECAT, no cats would. If there is

any dependence here, it is either symmetric or asymmetric in the wrong way to help Fodor's theory get this case right.

The cases involving more subtle differences between thinkers than the CAT/MAYBECAT case are more important to my argument, but they are unfortunately trickier to discuss. It is certainly true that they will not always have the same structure as these false belief cases, in which the thinker getting things 'wrong' is more restrictive in applying the relevant concept. And that structure is what makes it so clear (at least, so clear for a discussion of asymmetric dependence) that bringing in asymmetric dependence won't help Fodor account for conceptual publicity. But even so, many of the cases involving subtle differences will have that structure—many will be cases in which we intuitively want to disregard the way one or both of the thinkers is more restrictive in applying the relevant concept. And even ignoring this structure when it shows up, it is far from clear how a proponent of asymmetric dependence can consistently use it to achieve conceptual publicity in the subtler cases.

3.1.4 Deference does not solve the problem of publicity for Fodor

I mentioned in the last chapter that one of the few ways an individualistic theorist might try to account for A and B sharing CAT, if it is already stipulated that A possesses CAT, is by taking the features in virtue of which A possesses CAT into account when determining which concept B possesses. This general strategy, it seems to me, is Fodor's best hope when it comes to accounting for the publicity of concepts, and it does help him come closer to explaining some cases of concept sharing than he otherwise might. However, it does not take Fodor all the way to the publicity of concepts, at least if that publicity is to involve sharing which is anything like as widespread as we ordinarily take it to be.

One of the problems Fodor takes up in *The Elm and the Expert* is the problem of accounting for concept possession in the absence of full concept mastery. As the CAT/MAYBECAT case illustrated, this is closely related to the problem with which I am primarily concerned in this chapter and the last. In the standard case I have been considering, on a causal theory like Fodor's two thinkers will be 'prevented' from sharing CAT because they respond subtly differently to the many subtle variations among cats. In the CAT/MAYBECAT case, two thinkers are again prevented from sharing CAT because they respond differently to the variations among cats. The main differences between the cases are that in the CAT/MAYBECAT case the variations are much less subtle and we can clearly describe the source of the trouble as one of the thinkers falling short of mastering the concept.

As the title of the book suggests, the case Fodor focuses on in *The Elm and the Expert* is that of someone—Fodor himself—who can't tell an elm from a beech without the assistance of an expert:

> "I can't tell elms from beeches, so I defer to the experts." Compare: "I can't tell acids from bases, so I defer to the litmus paper"; or "I can't tell Tuesdays from Wednesdays, so I defer to the calendar." These three ways of putting the case are, I think, equally loopy, and for much the same reason. As a matter of fact, I

> *can* tell acids from bases; *I use the litmus test to do so.* And I can tell elms from beeches too. The way I do it is, I consult a botanist.
>
> . . .
>
> From the point of view of an informational semantics, the situation is *absolutely normal*: that my *elm* and *acid* thoughts have the content that they do depends on there being mechanisms that reliably correlate them with instantiations of elmhood and acidhood respectively. (1995, 34-35)

As I have discussed, Fodor holds roughly that the semantic value of a concept some individual possesses is determined by the causal laws in which uses of that concept feature as an effect. For example, Fodor would say that his thoughts involving the concept ELM are thoughts about elms in part because it is a law that *elmhood* causes him to use ELM. His point here, then, is that this is true both for a botanist who can recognize elms without help and for someone like Fodor who needs help. The only difference is that in Fodor's case, elmhood tends to cause uses of ELM by way of helpful experts to whom Fodor defers in his uses of ELM. But this is nothing special, since even for many experts on the subject *acidhood* tends to cause uses of ACID by way of litmus paper and other devices.

Fodor's point in the ELM case is simply that the mechanisms which sustain the correlation between uses of a concept and the property that that concept denotes can include other people, and that they do seem to include other people when a concept is used deferentially. We might expect this observation to help Fodor with the CAT/MAYBECAT case, then, if in that case B uses MAYBECAT deferentially. And more broadly, we might expect this to help Fodor overcome the arbitrariness of CAT if everyone coordinates in their use of CAT by deferring either to a single source or to multiple sources which all use CAT in the same way.

However, there is little reason to expect someone with false beliefs, like B in the CAT/MAYBECAT case[12], to reliably be more deferential than someone with true beliefs. Of course, B would probably defer to A in using MAYBECAT if B knew who was right and who was wrong about whether all cats have fur. But in most cases in which someone falls short of concept mastery by having a false belief, that person does not know that their belief is false—beliefs known to be false tend to be revised fairly quickly. B will likely be just as willing to use MAYBECAT without deferring as A will be to use CAT without deferring, and so Fodor will have a very hard time making deference relevant to this case.[13]

[12] As I have said, Fodor's theory would not count B as having false beliefs, but to discuss this case in neutral terms would be unnecessarily awkward here.

[13] Of course, B is more likely than A to eventually defer to some expert, since there is more in B's use of "cat" to correct than there is in A's. However, I can see no principled way for Fodor's theory to avoid treating this eventual change in behavior in response to correction by an expert as a case of conceptual change, rather than (as we would intuitively see it) as a case of B moving closer to mastery of an already-possessed concept. This is because B's eventual acquiescence to the expert's insistence that some cats do not have fur would change B's dispositions to use "cat," as well as the concept B expresses with "cat," and Fodor's theory would describe this change as a change in which concept B expresses with "cat"—perhaps

The difference between Fodor's ELM case and the CAT/MAYBECAT case is that in the ELM case Fodor's deference is total: experts are the only way he has to tell what an elm is, at least when beeches might be about. In the CAT/MAYBECAT case, on the other hand, B uses MAYBECAT confidently without assistance, perhaps deferring occasionally but not doing so as a rule. The question of whether deference can help Fodor overcome the arbitrariness of CAT, then, comes down to the question of whether the general case of using concepts in the neighborhood of CAT typically involves total deference, like Fodor's use of ELM, or involves only partial deference, like the CAT/MAYBECAT case.

Unfortunately for Fodor, I think it is fairly clear that deference is rarely if ever total, and that thinkers who fall short of mastery of a concept nonetheless frequently use that concept confidently without assistance. Whether the concept in question is CAT, RED, or KNOWS, many thinkers who have no special claim to expertise go on using the concept without waiting to be told how. This is, of course, quite reasonable: if we used all of our concepts as deferentially as Fodor uses ELM, we would never get any thinking done at all.

And even when thinkers do defer in using a concept, there is no guarantee that there is one monolithic source to which they will all defer. Focusing just on the CAT/MAYBECAT case, if we imagine the scenario in which A and B apparently disagree, it might be that they are both 'deferring' to different purported experts. And even with a subject like cats, on which there are some experts, the experts do not always agree. For example, experts on cats as a kind could disagree about whether the kind is a subgroup of a species or a clade, or about just which hybrids count as cats.

And even when thinkers do defer to a single source, they often will not be in a position to discover all the subtleties of that source's use of the concept, since many concepts are complex enough that they involve subtleties which are not frequently exhibited in behavior, as I mentioned in the last chapter when discussing the arbitrariness of CAT. The upshot, then, is that although our practice of deference might help Fodor account for concept sharing in a few interesting and important cases, we should not expect it to overcome the arbitrariness of CAT, or even the problem of false belief cases like CAT/MAYBECAT.

3.1.5 Other causal theories are no better off than Fodor's

Fodor's is not, of course, the only theory which assigns semantic values to concepts based on the mind-world causal relations in which those concepts stand or the information which is carried by use of those concepts. But the differences among theories in this family—though sometimes very significant—do not make a difference to my argument. This should be evident from my discussion above, since my argument took hold on the most basic version of Fodor's theory, which is fairly typical of theories in this family. The finer details of Fodor's theory, such as asymmetric dependence, on which Fodor differs from other theorists, only came in to the dialectic as potential responses to the argument on Fodor's behalf; my argument would

a change from expressing MAYBECAT with "cat" to expressing CAT with "cat"—because this change in dispositions would amount to a change in semantic value.

have applied to the theory even if it had been different with respect to these fine details, as long as it held on to its individualism.

For example, one theory that shares its causal core with Fodor's—though it is a theory of representation more generally, rather than a theory of conceptual representation[14]—is given by Dretske (1986). Both Fodor and Dretske treat a representation as representing what it does in virtue of the information it carries, which it carries in virtue of the causal relations in which it stands. Perhaps the most salient point of difference between Fodor and Dretske, though, is in how they handle cases which we would ordinarily think of as mistakes.

Many mistakes involve the use of a representation being caused by something which does not fall within the class denoted by that representation. For example, in the case I described above when discussing Fodor's notion of asymmetric dependence, a cat caused the use of the term "cow." On a theory which simply treats every kind of object which causes the use of a term as falling within the class denoted by that term, the upshot of a case like this would be that "cow" denotes (at least) the class of objects which are either cows or cats. Fodor brings in asymmetric dependence to solve this problem, but Dretske brings in something very different: the notion of a biological function.

The full motivations for—and the full development of—Dretske's theory are outside the scope of this paper, but in order to further justify my claim that my argument is not specific to Fodor's theory, I want to show how this primary difference between Fodor's and Dretske's theories does not make Dretske's theory immune from my argument. The relevant point about Dretske's theory, then, is that Dretske holds that a system's being in a certain state represents the external states which cause that system to be in that state when the system is not malfunctioning.

For example, someone applying Dretske's theory to the case of the concept CAT might say that CAT being deployed in a thinker's conceptual system somehow represents cats because when the conceptual system is performing its function properly, only cats cause the conceptual system to deploy CAT.[15] The extra resources that might help Dretske overcome the arbitrariness of CAT, then, have to do with the function of a thinker's conceptual system.

It seems to me that the most straightforward way someone with a theory like Dretske's might try to use the special resources their theory offers to overcome the arbitrariness of CAT is by claiming that it is part of the function of all thinkers' conceptual systems to indicate the state of the world with respect to cats by deploying CAT. The upshot of this would be that all thinkers would be able to use the concept CAT, thereby denoting cats,

[14]One other way in which Dretske's theory, taken as a whole, is not relevant here is that Dretske's theory of representation is not individualistic, at least for all representations. However, because Dretske does not make clear how the theory should be applied to conceptual representations, it is left open whether the theory is individualistic or not when it comes to conceptual representations. As I will show below, it seems to me that one could apply the theory to conceptual representations in an individualistic way or in a more social way, and the differences in these applications tell in favor of the claim I am trying to support here.

[15]This is probably not the most faithful way to apply Dretske's theory to this case, since I have put it in terms that make comparison to Fodor's theory easier than it really ought to be. However, the application should still be faithful enough for my purposes.

regardless of how any particular thinker's uses of CAT are caused, because the denotation of CAT is a part of the very function of all thinkers' conceptual systems.

However, it is far from clear how someone working with a theory like Dretske's could motivate the claim that our conceptual systems have such a specific function. The functions that Dretske would use here are biological functions, given to a system in part by evolution[16] , and it is hard to see how our evolutionary history could be so specific in assigning a function to our conceptual systems as to include in the function a role for CAT—not to mention concepts with a shorter history, like CARBURETOR. This straightforward way of bringing in functions in an attempt to solve the problem of publicity, then, will not work.

The other option for bringing Dretske's resources to bear on the problem of publicity is to claim that publicity of concepts in general is somehow built in to the function of our conceptual systems, even if the publicity of particular concepts like CAT is not built in because those concepts are too evolutionarily unimportant or recent. For example, one might claim that generally, the function of a thinker's conceptual system in deploying a concept C is to indicate the state of the external world with respect to whatever causes the use of C by anyone who possesses C—or something along these lines.

This is, I think, a very promising line for someone with a theory like Dretske's to take, both because it seems independently plausible (if a bit simple) and because it has a real hope of overcoming the arbitrariness of CAT and other concepts like it. However, I wish to highlight one very important feature of the claim: in order to use a theory involving this claim to determine the semantic value of a concept C, one must first determine to whom "anyone who possesses C" applies, because the uses of C by anyone who possesses it are potentially relevant to C's semantic value. But determining who possesses C amounts to determining who shares C, so a theory involving this claim is not an individualistic theory of concepts, but is rather a social theory of concepts. Taking the question of who shares a concept to be prior to the question of (say) that concept's semantic value is at the core of social theories of concepts, as a family of theories. It is no surprise, then, that I find this line promising, since I am trying to motivate social theories of concepts as a family, but this is not a line that an individualistic theorist can take without giving up their individualism.

The upshot of this section, then, is that causal theories of concepts, such as Fodor's theory or a theory that determined the semantic values of concepts using Dretske's machinery, cannot account for the publicity of concepts while remaining individualistic theories. In light of my argument from the last chapter, this should lead us to conclude that on individualistic causal theories of concepts, we should expect that concepts (at least complex general concepts like CAT) are not public. And given the assumption that publicity of concepts is an

[16]Dretske's account of functions does bring in associative learning (1986, section 5), but only to solve the problem of evolution not giving determinate-enough functions on its own. Evolution still has to do most of the heavy lifting in giving systems their biological functions on Dretske's picture. Also, even if someone with a theory like Dretske's does try to assign a significant role to associative learning, that will not help to solve the problem of publicity, because there will be many subtle differences among the learning histories of different thinkers with respect to cats, just as there are many subtle differences among the features of thinkers with respect to cats that Fodor's theory takes to determine semantic values.

important desideratum for a theory of concepts, this should lead us to favor other theories over individualistic causal theories of concepts.

Just to be clear, though, I should reiterate that I think it is possible to have a causal theory of concepts without having an individualistic theory of concepts. My complaints in this section are specific to individualistic theories and do not apply to causal theories more generally. Indeed, I myself am partial to determining the semantic value of a concept based on the causal relations in which uses of that concept stand. I just think this should only be done after determining which thinkers share that concept—that is, I think this semantic theory should be embedded in a social theory of concepts, rather than in an individualistic theory.

3.2 Inferential-role theories of concepts

Having dealt with causal theories of concepts, I wish now to move on to a family of theories according to which a concept is individuated[17]—and its denotation, if there is such a thing, is determined[18]—at least in part according to its role in the thought of an individual who possesses it. Theories in this family differ as to which aspects of a concept's role are individuating, but they typically include inferences of a certain form involving the concept that an individual is disposed to make (or is disposed to make in a certain way), perhaps along with intentions and perceptual judgments[19] of a certain form involving the concept which an individual is disposed to form (or is disposed to form in a certain way). This description might be too high-level to make much sense now, but things should become more clear when I discuss some examples below.

[17]Some theories that assign a critical role to inferential or conceptual role nevertheless do not take these features as they show up in an individual's thinking to be of primary importance in answering questions about concept individuation, or at least questions about concept sharing. For example, the theory Robert Brandom develops in *Making It Explicit* (1994) identifies a concept with its role in inference, very broadly construed, but does not take the question of whether two thinkers share a concept to depend on the question of whether a concept plays the same role in one thinker's inferences that a concept plays in the other thinker's inferences. Rather, Brandom is more interested in the role a concept plays in the inferences of all thinkers who share the concept, and he answers questions about concept sharing by other means. In other words, Brandom has a social theory of concepts.

[18]I do not wish to discuss the question of whether concepts importantly have denotations. My discussion will not primarily focus on views that make no mention of concepts' denotations, that deny that concepts importantly have denotations, or that involve the claim that the meaning of a concept just is the concept's role in thought. However, my reasons for focusing on views involving denotations are mostly presentational: my discussion of causal theories of conceptual content is necessarily concerned with denotations, and so presentation can be simplified elsewhere if I put this discussion in similar terms. Also, I am less interested in differences in conceptual role that do not make for a difference in denotation, since these differences might not seem to threaten publicity as gravely. I hope it will be clear enough that most of the important things I have to say still apply to views that somehow individuate concepts according to their role in an individual's thought but assign no importance to denotation.

[19]Attitudes other than intention and perception might be relevant as well, depending on the theory, but I will focus on these in addition to inference.

In the first chapter of this work, I discussed the case of a theory according to which a concept is individuated by its total role in the inferences of a thinker who possesses it. The upshot of that discussion was that this theory leaves no room for publicity, since there is no reason to expect two different thinkers to make all the same particular inferences involving a concept, and so on this theory there is no reason to expect two different thinkers to share a concept. This theory is a member of inferential-role family of theories, but it is an extreme example, and there is little reason to expect a proponent of such a theory to be concerned with the publicity of concepts (in the present sense), since the theory clearly leaves no room for it.[20] Other theories somehow restrict the aspects of a concept's role that are individuating, and since these theories might seem to leave more room for the publicity of concepts, they are of greater interest in the present discussion.

There are different ways to restrict the individuating aspects of a concept's role, but it is fairly typical to focus on inferences, perceptual judgments, and intentions which all have some special feature, such as being specially compelling, making up part of a special kind of theory, or seeming specially appropriate to the thinker in question. Among these special inferences, the focus might be on certain formal features of the thought(s) inferred from and the thought(s) inferred to, whereas among the special perceptual judgments, the focus might be on certain aspects of the experience or of the scene perceived, and among the special intentions, the focus might be on certain aspects of the act which results if the intention is carried out.

3.2.1 Peacocke has an inferential-role theory

The particular view I will focus on as an exemplar of views in this family is laid out by Peacocke in *A Study of Concepts*, and it treats only a greatly restricted part of a concept's total inferential role as individuating: the inferences that are "primitively compelling." For example, Peacocke identifies the concept AND as—very roughly—the concept such that anyone who possesses the concept finds primitively compelling all and only inferences involving the concept that take the forms P AND Q, therefore P; P AND Q, therefore Q; and P, Q, therefore P AND Q. (1995, 6)

Peacocke tells a somewhat different story about concepts that do not primarily feature in inferences from one thought to another. For example, he identifies the concept RED as—leaving out some details that are very important but not relevant to my argument—the concept such that anyone who possesses the concept, on having an experience that (in a certain nonconceptual, salient way) presents an object as red, is disposed to form the judgment THAT IS RED (on certain grounds and not others). (1995, 7) These are the kinds of things that Peacocke identifies as the individuating roles of concepts: certain mental transitions involving a concept, such as the primitively compelling inferences and

[20]Some individualistic holistic inferential-role theorists, such as Block (1986), do try to account for something like publicity of concepts by focusing on similarity, rather than sameness, of concepts. I see little hope for this kind of move, but the move to similarity in general is the focus of the next chapter, so I will return to holism then.

the perceptual judgments formed on the right kind of grounds, are part of the individuating roles, and other mental transitions are not part of the individuating roles.

3.2.2 A concept's semantic value depends on its role in thought

Finally, in some inferential-role theories, once a concept's inferential role has been identified, its semantic value is determined with an eye to preserving truth in the relevant inferences, accurately representing the perceived world in the relevant perceptual judgments, and acting successfully when acting from the relevant intentions.[21]

For example, if BACHELOR is identified by its inferential relations to UNMARRIED and MAN, an inferential-role theory might give the verdict that it denotes *unmarried-man-hood*, since BACHELOR's having that denotation allows those inferences to preserve truth. Of course, this assumes that the theory can account for UNMARRIED and MAN having their expected denotations, and this cannot simply be taken for granted. If a theory identifies and assigns semantic values to concepts based only on relations among concepts, there is a significant danger of the theory giving explanations that never 'bottom out.' If a theory holds that the semantic value of BACHELOR depends on the semantic value of MAN (among other things), and the semantic value of MAN depends on the semantic value of HUMAN (among other things), and the semantic value of HUMAN depends on the semantic value of ANIMAL (among other things), then one might reasonably wonder whether all these dependencies will ultimately be resolved if the theory's explanations will always involve further concepts.

This danger is one of the reasons that many inferential-role theories assign just as much importance to concepts' relations to perception and action as to their relations to other concepts, both in identifying concepts and in assigning them semantic values, since this allows explanations that reach outside of a conceptual web to phenomena in the world.[22] So, for example, if RED is identified by its relation to experiences which present objects as red, then (continuing with my great oversimplification of the analysis of perceptual concepts) since experiences typically accurately present objects, the theory might assign the concept the denotation of *redness*. If SIT is identified by its relation to intentions which result in sitting, then it might be assigned the denotation *sitting* (or *sittinghood*).

[21] In *A Study of Concepts*, Peacocke focuses on concepts' place in judgments, perceptual or otherwise, more than on their place in intentions, but he advocates a theory which clearly fits this model, in that it assigns semantic values with an eye to the correctness of the practices by which a concept is identified:

> The [theory of how the semantic value of a given concept is determined from its possession conditions] (together with the world in empirical cases) assigns semantic values in such a way that the belief-forming practices mentioned in the concept's possession conditions are correct. (1995, 19)

[22] This is one of the main points Gilbert Harman makes in "Conceptual Role Semantics" (1982).

3.2.2.1 Inferential role semantics

It is perhaps worth repeating here that many theories of concepts that fall into the inferential role family, as I have described it, do not assign concepts any semantic values—or do not assign any importance to concepts' semantic values—beyond their roles in thought. Indeed, one of the reasons that I am focusing my discussion on Peacocke's theory is that unlike many other theorists[23], he does discuss semantic values that are distinct from inferential roles, and this discussion of his allows greater symmetry between my discussions of his and Fodor's theories.

However, whether an inferential-role theory does or does not assign separate semantic values to concepts makes little difference to my argument. The feature of these theories that my argument will rely upon is just that they individuate concepts according to at least part of their roles in the thoughts of individuals.

The only way that semantic values like denotations are relevant to this part of my argument is that for theorists like Peacocke who make use of them, they make room for a potential objection that I will take care to block: If an inferential-role theory is incompatible with publicity because different thinkers rarely have concepts with the same individuating inferential role, a theorist like Peacocke might claim that this is not a serious problem, because what really matters is that different thinkers commonly have concepts with the same semantic value. So in the argument that follows in this section, when discussing differences in individuating inferential roles I will make sure to focus on differences in inferential roles that are reflected in differences in the semantic values those inferential roles determine. I hope it will be clear, though, that although this focus will enable my argument to apply to theories like Peacocke's, it will still apply just as well to other inferential-role theories that involve different metasemantic views.

3.2.3 Peacocke's theory shares some problems with causal theories

Since perceptual concepts like RED are identified in part by their relation to certain perceptual states, they will be individuated by Peacocke's theory—and by other inferential-role theories which are like Peacocke's in this respect—in much the same way that causal theories like Fodor's individuate concepts. This is because perceptual states are typically individuated at least in part according to causal relations (either causal relations involving them or causal relations they involve), so if a concept is individuated at least in part according to which perceptual states it is tied to in inference (or something like inference), it will end up being individuated at least in part according to mind-world causal relations very like those involved in Fodor's account.

For example, if to be the concept RED is in part to be a concept which is connected in a certain way to red perceptions, and to be a red perception is in part to involve or be involved

[23]For example, Harman (1982) and Brandom (1994) offer theories on which concepts do not importantly have semantic values that go beyond their roles in thought.

in a certain causal connection to the property *redness*, then RED will not be very widely shared. Given the subtle variations in color sensitivity among human visual systems, for many people who possess a concept which they express using the word "red," their concept will be connected to some color property which is very similar but not identical to *redness* in just the same way that RED is connected to *redness*. The upshot is that, as with Fodor's theory, on Peacocke's theory we should expect to find, rather than a single concept RED which denotes *redness* and which is shared by many people, many concepts RED_1, RED_2, and so on which respectively denote many properties $\textit{redness}_1$, $\textit{redness}_2$, and so on and which respectively are shared by few people or only one person.

And if Peacocke identifies less perceptual and non-perceptual concepts according to their inferential relations—direct or indirect—to perceptual concepts, this problem of publicity spreads from the perceptual concepts to the less perceptual and non-perceptual concepts. For example, it might be that to possess the concept CHERRY TREE—surely not a perceptual concept in the way RED is, since at least for many of us our experience does not present cherry trees as such—one must find certain inferences involving RED primitively compelling. These inferences might include inferences from a thought of the form THAT IS A CHERRY TREE to a thought of the form THAT TYPICALLY PRODUCES RED FRUIT (or something a bit more nuanced). If one must find inferences like this primitively compelling in order to possess the concept CHERRY TREE, a thinker who possesses RED_1 rather than RED would not be able to possess CHERRY TREE, but rather could only possess some similar concept like CHERRY TREE_1. And the same line of reasoning could then be applied to any concepts which depend on CHERRY TREE, and so on.

In the end, Peacocke will have the same kind of trouble with any even slightly empirical concepts that Fodor has with all the concepts to which his theory applies. And since a great many concepts are empirical in the relevant sense, Peacocke will have a great deal of trouble accounting for the publicity of concepts if it is to include anything like all the concepts we ordinarily take it to include.

That being said, in the rest of this section I want to focus on problems—and potential solutions—that are specific to inferential-role theories like Peacocke's, rather than on the problems they have in common with causal theories. Since the heart of inferential-role theories is the identification of a given concept according to the part of the concept's role in inference (and possibly other mental transitions) that individuates that concept, the problem of publicity that is specific to these theories arises thanks to arbitrariness in this individuating role.

3.2.4 There is arbitrariness in the individuating role for CAT

Let us now return to our running example, the concept CAT. It isn't immediately obvious what CAT's individuating conceptual role should be, but it might involve inferences of the form *X* IS A CAT, therefore *X* IS A MAMMAL, IS NOT A DOG, *et cetera*, AND IS THE KIND OF THING WHICH TYPICALLY MEOWS, HAS FUR, GROWS INTO THIS SHAPE (identifying a certain cat-like shape), CAN INTERBREED WITH OTHER CATS,

et cetera. Or perhaps the truly important inferences are more along the lines of *X* IS A CAT, therefore *X* HAS THE PROPERTY THAT THE THINGS AROUND HERE THAT TYPICALLY MEOW, HAVE FUR, GROW INTO THIS (cat-like) SHAPE, *et cetera* HAVE AND THAT BEST EXPLAINS WHY THEY TYPICALLY DO THOSE THINGS.

There is, of course, a fairly famous problem in claiming that there is a single pattern of inference that is precise enough to distinguish a concept like CAT from all other concepts (including concepts with very similar meanings, like CAT OR WILDCAT and NATURALLY FURRY CAT) and yet imprecise enough that everyone who possesses the concept follows the pattern in using that concept. The problem is that, even among philosophers who spend a great deal of time analyzing concepts like CAT, there has very rarely been anything like consensus about what a particular concept's pattern should be. This suggests that there might be no single set of inferences that are the individuating inferences for CAT, which is a serious potential problem for inferential role theorists. However, this is not the problem I want to press. Instead, I want to grant the inferential-role theorist the claim that CAT has a single individuating inferential role, and I want to claim that even in this single individuating role, there is room for arbitrariness.[24]

The individuating inferences for CAT, whatever they are, will have to be fairly numerous or complex to distinguish cats from dogs, wildcats, and other similar animals. Or they will have to involve other fairly complex concepts, such as the concept mentioned above as THIS SHAPE, which would be a (perhaps perceptual) concept of the typical shape which adult cats grow into. And, as became clear in the last chapter, with complexity comes arbitrariness. For example, the requirements in THIS SHAPE might be loosened or tightened slightly so as to allow more or fewer cat-hybrids and cat-ancestors to have the denoted shape without clearly making the resulting CAT concept less useful. Similarly, there is no clearly decisive reason to include or to exclude a connection to the concept FUR (or MEOWS, or DOG) among the individuating inferences for CAT.

And this arbitrariness in the individuating role for CAT is not merely the arbitrariness of selecting from different conceptual roles which are all equivalent when it comes to determining the denotation of CAT. If *cathood* amounts to some complex genetic property, then it should be fairly clear how at least some of the arbitrariness involved in identifying which inferences are individuating is reflected in the arbitrariness of identifying which complex genetic property *cathood* amounts to. If the inferences involve any concepts of observable attributes which are typical of cats but very similar to attributes found in cat-hybrids or non-cat cat-ancestors, and which can be explained by genetic properties which are typical of cats but very similar to genetic properties of cat-hybrids or non-cat cat-ancestors, then the two kinds of arbitrariness will be fairly directly linked.

[24]Of course, I think these two phenomena are related: one of the main reasons that it seems impossible to identify the individuating role of a concept as complex as CAT is that the concept involves a great deal of arbitrariness, and to identify an individuating role one would have to either do so in a precise but arbitrary way—which would leave one open to criticism from others who have identified different but equally satisfactory individuating roles—or do so in an imprecise way—which would leave one open to criticism from others who are not satisfied with such imprecision.

For example, if THIS SHAPE, in identifying a cat-like shape, includes requirements on ear length, and some cat-hybrids or cat-ancestors have ears which are very slightly too long to meet the requirements, and this slight variation in ear length can be explained by slight genetic variation, then the arbitrariness with respect to requirements on ear length in selecting the concept THIS SHAPE to be involved in the individuating inferences for CAT will be reflected in the arbitrariness in selecting the genetic properties responsible for ear length to be part of the complex genetic property which *cathood* amounts to.

On the other hand, if *cathood* amounts to some ancestral property, the arbitrariness might come in somewhere other than the inferences involving more perceptual concepts like THIS SHAPE. If CAT denotes an ancestral property, then this should somehow be reflected in the individuating inferences for CAT—perhaps because the individuating inferences require that *cathood* be the property that allows a certain kind of explanation of why cats typically exhibit the outward symptoms by which we initially identify cats, such as meowing, having fur, and growing into a certain shape, and it turns out that the property that allows this kind of explanation is a certain ancestral property.

Now, there are many ancestral properties that could be cited in explaining why some animals exhibit the outward symptoms by which we initially identify cats, and some of these properties are not good candidates for being *cathood*. If we held that the individuating inferential role of CAT required that *cathood* give a certain kind of explanation, and we wanted to get more clear about what this inferential role is and what requirements it puts on this explanation, we might take up these bad candidate ancestral properties and figure out how, if at all, the explanation each gives fails to meet our standards.

For example, one could explain why all currently living cats exhibit the outward symptoms of *cathood* by pointing to the fact that they are descended from a certain population of animals that exhibited similar symptoms, a population that happens to include all and only parents of currently living cats. But *cathood* surely does not amount to being descended from that population—for one thing, many cats were born before any animal in that population was alive. The problem with this ancestral property and other silly ancestral properties like it is plausibly that they do not give the best explanation of why currently living cats exhibit the outward symptoms by which we initially identify cats. The problem is not that they do not give any explanation of this, since they clearly can—the explanations they give are just bad explanations, because the ancestral properties they involve are too gerrymandered. To rule out such silly ancestral properties from being *cathood*, then, the individuating inferential role will ultimately have to give us some standards to rule out gerrymandering—standards for geographic or temporal contiguity or clustering. But, as came out in the last chapter, standards like these bring in a great deal of arbitrariness, and this arbitrariness can be reflected in the way CAT is related in inference to concepts like CONTIGUOUS or CLUSTERED. So, even if *cathood* amounts to an ancestral property, there is arbitrariness in the individuating inferential role of CAT.

Unfortunately, I am not able to list all the potential individuating roles for CAT and show where arbitrariness comes in for each one. As I mentioned above, it is far from easy to figure out what the individuating role for CAT should be. But I hope the examples I

have discussed have sufficiently illustrated the complexity which must be involved in any individuating role for a concept like CAT, as well as the great potential for arbitrariness to come in to any individuating role of sufficient complexity.

Similarly, I am not able to give examples of plausible individuating roles for all the kinds of concepts which might be relevant to my argument. And there could surely be some individuating roles which do not involve the kind of arbitrariness that CAT's does—the example of AND, discussed at the beginning of this section, shows that some concepts might have very straightforward individuating roles. However, once we leave the realm of logical and mathematical concepts, we should expect most of the concepts we consider to have individuating roles more like CAT's than like AND's. It would be very difficult indeed for an inferential-role theorist to account for the publicity of concepts just by focusing on simple concepts like AND.

3.2.5 Where there is arbitrariness, publicity needs explanation

Since there is arbitrariness involved in specifying the individuating role for CAT, there are many subtly different individuating roles which would be just as useful to us as CAT's role. And since each such role individuates a different concept, there are many subtly different concepts which would be just as useful to us as CAT, many of which (as discussed above) would have subtly different denotations.

But if there are many subtly different concepts which would be just as useful to us as CAT, and many of these differences are so subtle as to be undetectable in ordinary practice, and which concept an individual possesses is just a matter of which role a concept plays in her own thought, then why should we expect many people to possess the same concept?

For example, suppose one of the individuating inferences for CAT involves a concept of a certain cat-like shape. Then, for the many subtly different concepts of certain cat-like shapes, there will be many concepts CAT_1, CAT_2, *et cetera*, all of which are subtly different from CAT and from each other: someone who finds a certain bunch of inferences involving THIS SHAPE_1 primitively compelling will possess CAT_1, *et cetera*. So if different people tend to use subtly different cat-like-shape concepts in their inferences (And why shouldn't they, if all these concepts serve as well as one another for their purposes?), then they will tend not to share the concept CAT but instead to have subtly different concepts like CAT_1 and CAT_2.

An inferential-role theorist who wishes to account for the publicity of CAT, then, must do something to explain how thinkers manage to coordinate on the concept THIS SHAPE and coordinate in connecting it to CAT in their inferences, as well as how these thinkers manage to overcome all of the other arbitrariness that CAT involves.

Luckily, Peacocke is aware of a problem very closely connected to this one, and he offers a potential solution. Unluckily, it is not entirely clear how his solution is meant to work or whether it is available to a theorist who wishes to have an individualistic theory of concepts.

3.2.6 Peacocke and belief attribution without concept mastery

Just as Fodor is aware of the potential for his theory to have a problem with publicity, so is Peacocke. But whereas Fodor tries to show that his theory already possesses the machinery necessary to solve the problem, as I discussed above, Peacocke tries to solve the problem by offering an amendment to his theory. As I will explain, although I am somewhat sympathetic to Peacocke's offering, it is missing some important details that need to be settled before it can be considered as a potential solution to the problem of publicity—and it is far from clear how these details could be settled on any individualistic theory of concepts.

As I explained above, Peacocke is concerned with concepts' possession conditions. That is, he is concerned conditions that thinkers can meet or fail to meet such that possessing a given concept just is meeting that concept's condition. And when discussing perceptual concepts, he proposes a possession condition for the concept *red*[25] that implies that a thinker who "is properly perceiving a book of a dark red color and is taking his perceptual experiences at face value" must be disposed to believe something like *that's red* (1995, 7 and 28). But Peacocke goes on to consider a puzzle that seems, at least at first, to put some pressure on anyone with a theory of possession conditions like his:

> It can be introduced by the following four propositions. They are all superficially plausible, but they are jointly inconsistent. To fix ideas, I take the concept *red* as an illustration. Many other concepts, with appropriate alterations, could be made to serve in its place.
>
> (1) The above possession condition for *red* ... is correct.
>
> (2) The construction "x believes that ____ red ____[”] expresses a propositional-attitude relation to a structured thought containing the concept *red* as a constituent.
>
> (3) A thinker who has attitudes to thoughts containing the concept *red* must fulfill its possession condition.
>
> (4) A thinker who has false beliefs about the boundaries of the color spectrum covered by the word "red" and who sincerely says of an object "It's red" or "It's not red" is correctly described as believing the object in question to be red, or not red, respectively. (Peacocke, 1995, 27-28)

(1) through (4) are inconsistent because it follows from (2) that the thinker described by (4) has an attitude to a thought containing *red*, and it follows from that and (3) that the thinker must fulfill *red*'s possession condition, and it follows from that and (1) the thinker must be disposed—at least in the proper conditions, as described above—to judge any red thing to be red, but by (4) that thinker has false beliefs about the boundaries covered by "red" and

[25] For this subsection, I will adopt Peacocke's style of using italicized words, rather than capitalized words, to name concepts.

so is presumably disposed—even in the proper conditions—to judge some red things to be non-red.

Peacocke solves the puzzle by rejecting proposition (3):

> Premise (3) is false because possession conditions are different from attribution conditions. A possession condition states what is required for full mastery of a particular concept. The attribution conditions for *red*, the conditions under which something of the form "*x* believes that ____ red ____"is true, are much weaker than the possession condition. The following are jointly sufficient for such an attribution to be true.
>
> (a) The subject is willing sincerely to assert some sentence of the form "____ red ____"containing the word "red"(or some translation of it).
>
> (b) He has some minimal knowledge of the kind of reference it has (e.g., that it is a color word).
>
> (c) He defers in his use of the word to members of his linguistic community. (1995, 29)

I do find Peacocke's puzzle compelling, since I find (1) through (4) quite plausible and yet I cannot accept them all. In a way, the positive parts of my project are meant to solve precisely this puzzle, or some more general version of it. And I am sympathetic to solutions which bring in public language in an important way, as Peacocke's does. However, it seems to me that in offering such a quick solution, Peacocke might be failing to appreciate the true scope of the problem.

As I argued in the last chapter, on individualistic theories of concept possession like Peacocke's, 'full mastery' of concepts like *red* (as well as *cat*, *tree*, *chair*, and most other concepts of significant complexity) will be quite rare, achieved only with a great deal of luck. Instead of having many people mastering a single concept like *red*, these individualistic theories will have many subtly different concepts (like red_1, red_2, et cetera) each being mastered by only one or a few people.

So if, on Peacocke's theory, many people are to have *red*-involving beliefs, it will only be because the vast majority of them meet the attribution conditions for *red*, rather than the possession conditions. At the very least, this motivates more thoroughly spelling out what deference amounts to—which is no easy feat, as came out in my discussion of Fodor's use of deference—as well as what is required to be able to use the word "red" or some translation of it, especially since people who do not possess *red* will need to be able to use "red" if attribution conditions aren't to be stronger than possession conditions. The solution I will ultimately propose for this kind of problem will importantly involve both of these notions, and it will turn out to be fairly tricky to make good use of these notions without already having some other solution to the problem of publicity.

One way to see why it is important to make clear the details of Peacocke's proposed solution is to consider the following challenge: Someone who, as in proposition (4) above,

has false beliefs about the boundaries covered by "red" is typically disposed to defer to members of his linguistic community not because he knows he has false beliefs or because he is less certain in his beliefs than someone with true beliefs, but rather because in general we are all disposed to defer to members of our linguistic communities, whether we are right or not. So someone who has, on Peacocke's theory, mastered the concept *red* will typically be just as likely to meet conditions (b) and (c) for attribution of *red* as someone who has not mastered the concept. But conditions (b) and (c) for attribution of *red* are no different from the analogous conditions for attribution of some possible alternative concept, red_1, which covers a portion of the spectrum which exactly corresponds to the portion of the spectrum which some particular subject described by proposition (4) believes "red" covers. We have, then, two ways of describing a situation in which two people, A and B, differ on the boundaries covered by "red":

A is right: A possesses the concept *red*, but B fails to do so thanks to having false beliefs about the boundaries covered by *red*. B is able to have false beliefs involving the concept only because B meets the attribution condition for *red*, thanks in part to B's ability to use the word "red" which is somehow closely tied to *red*.

B is right: B possesses the concept red_1, but A fails to do so thanks to having false beliefs about the boundaries covered by red_1. A is able to have false beliefs involving the concept only because A meets the attribution condition for red_1, thanks in part to A's ability to use the word "red" which is somehow closely tied to red_1.

If only one of these descriptions is correct, in virtue of what is it correct? If both are correct, how are we to make sense of that?

This challenge is meant not as an objection, but rather as a request for further explanation, particularly of the relationship between "red" and *red*, which seems to me to be the only reasonable way of differentiating the two descriptions.[26] Because Peacocke is putting so much theoretical weight on this relationship, it deserves more attention.

In light of the argument I have made over these last two chapters, it should be clear that an individualistic theorist will not be able to easily explain the relationship between "red" and *red* by claiming that most people use the word "red" to express the concept *red* which they possess, since (unless this theorist already has some other solution to the problem of publicity) that theorist is not in a position to claim that most people possess the concept *red*. And other easy explanations will be similarly out of reach for individualistic theorists—indeed, as far as I can tell there is no prospect for a theorist making use of this relationship to solve the problem of publicity unless that theorist gives up on individualism.

Lastly, Peacocke's proposed solution also requires a more detailed explanation of the relationship between being able to have beliefs involving a concept and possessing that concept. Peacocke seems to me to be entirely correct in distinguishing between the case of

[26]The most obvious other way of differentiating the two would be to somehow deny the existence (or accessibility, at least to humans) of red_1 while affirming the existence (or accessibility) of *red*. I discussed the problems with this kind of move in the last chapter in the subsection on naturalness and eligibility.

fully mastering a concept and the case of being able to have beliefs involving that concept. I am simply more interested in the latter case, since for most concepts of significant complexity, we will only ever be able to count on being in the latter case[27]—for most concepts of interest, full mastery will only come about with a great deal of luck. It is therefore strange to me that Peacocke only calls cases of full mastery cases in which a thinker 'possesses' a concept, since they are not the ordinary cases, and in discussions of concepts we usually call the ordinary relationship between thinkers and concepts 'possession.'

3.2.7 Other inferential-role theories are relevantly similar

As I have described it, all that is required for a theory of concepts to belong in the inferential-role family is that the theory individuate concepts according to all or part of their role in an individual's perception, thought, and action.[28] Since this clearly leaves room for a great deal of variation, I have some reason to worry that my argument that Peacocke's theory has a problem with publicity might not apply just as well to other theories in this family.

However, the specific features of Peacocke's theory that distinguish it from other theories in this family are simply not relevant to the main thrust of my argument. This should be clear upon review of the argument, since it does not crucially involve reference to any of the distinguishing features of Peacocke's theory, except for the theory's making a distinction between a concept's total role and its individuating role.

This feature of Peacocke's theory distinguishes it from theories that individuate concepts according to their total role in the inferences of an individual (or perhaps their total role in an individual's inference, perception, and action), so the argument I made above does not directly apply to these individualistic holistic conceptual-role theories. But I have already discussed the very serious problem these theories have with publicity, so this is not a problem for my argument.

---

[27] As will become clear in chapter 5, on a social theory there will be little importance attached to the question of whether a given individual has fully mastered a given concept, since on a social theory concept possession will be treated as a kind of community membership. However, if conditions of mastery can still individuate concepts on a social theory, there will be a sense in which a community must 'master' the concept its members possess.

[28] With this criterion, I am identifying a slightly different family of theories than the families that are often called "conceptual role semantics" (Harman, 1982) and "inferential role semantics" (Boghossian, 1994). The main difference between these families of semantic theories and the family I am interested in is that these families identify the meanings or contents of concepts with all or part of their conceptual or inferential role and (often implicitly) individuate concepts according to their meanings, whereas the theories I am really interested in need make no mention of meanings or contents—I am concerned just with individuation. (Another difference is that some of the theories in these semantic families primarily concern public-language words in addition to or instead of concepts, whereas in this chapter I am concerned just with concepts.) But the theories in these semantic families face the same important decisions as theories in the family in which I am interested: for example, in discussing inferential role semantics, Paul Boghossian draws particular attention to the question of how a "meaning-constituting inferential role" is to be constructed out of a "total inferential role" (1994, 110), which is one of the central questions Peacocke's theory gives an answer to.

There is, however, one distinguishing feature of Peacocke's theory that is relevant to my argument, though its relevance falls outside what might be called the argument's 'main thrust,' since it is intended as a response to an argument like mine. This feature is Peacocke's claim that we must distinguish between the two kinds of access a thinker could have to a concept—access by meeting the concept's possession conditions and access by merely meeting the concept's attribution conditions—and use this distinction to solve a problem closely related to the problem of publicity. Now, because this feature of the theory is intended as a response to an argument like mine, it deserved and received particular attention. So one might reasonably wonder whether other inferential-role theories might feature other attempted solutions to the problem of publicity and whether my argument might not cover these intended solutions.

Of course, I cannot anticipate all attempts at solving the problem of publicity. But all of the attempts I have seen—at least, all of the attempts that are compatible with individualism in a theory of concepts—follow along the same lines as either Peacocke's or Fodor's attempts.[29] The only important thing that distinguishes Peacocke and Fodor is that they offer the best-worked-out attempts of their respective kinds.

So when it comes to individualistic non-holistic conceptual-role theories—theories that, like Peacocke's, individuate concepts according to parts of their roles in thought, perception, and action—my argument should apply even if the theories are somehow different from Peacocke's. However a theory tries to identify concepts' individuating roles, when it comes to a complex concept like CAT there will be enough arbitrariness in the concept's individuating

[29]For example, Harman attempts a solution to a related problem that has some features in common with both Peacocke's and Fodor's attempts. Harman is discussing the case of someone who somehow possesses the concept of an oak tree but has not mastered the concept—e.g. cannot distinguish oaks from elms. He proposes that the content of their concept derives from its relation to the public-language term "oak" as used by an expert:

> Of course, ... there are other people who can recognize oaks and distinguish them from elms and who know various distinguishing properties of the trees. These other people may have a concept of an oak tree which has functional roles that are sufficient to make it the concept of an oak tree apart from any relations the concept has to the word "oak". It is plausible
>
> 1. that their concept acquires its content from this aspect of its functional role, i.e., its role apart from its relation to the word "oak"
> 2. that the word "oak" as they use it has the content it has because of its connection with their concept of an oak tree
> 3. that the word "oak" as used by a more ignorant person can have the same content by virtue of connections between that person's ignorant use of the word and the expert's use
> 4. that the content of the more ignorant person's concept of an oak tree derives from its connection to his or her use of the word. (Harman, 1982, 248)

Harman's attempted solution depends on the dubious expert/ignorant distinction more than Peacocke's does, and it isn't as well-integrated in to the surrounding theory as Fodor's is. It leaves if anything more unanswered questions than Peacocke's or Fodor's attempts, but it itself is a more detailed working out of a proposal from Putnam (1975).

role that we have no reason to expect different individuals both to have concepts with that particular role, rather than having concepts with merely similar roles. Thus, the theory will have a problem with publicity.

3.3 Conclusion

So both causal and inferential-role theories of concepts have a problem with publicity, at least if they are individualistic theories—theories that identify the concept an individual possesses prior to determining who else shares that concept. And given that other individualistic theories face very similar challenges to those faced by causal and inferential-role theories—challenges that include identifying complex concepts like CAT based only on individual possession conditions while still distinguishing them from other similar complex concepts—they will likely end up in the same boat. Even though I have not been able to discuss other families of individualistic theories of concepts in as much detail as causal and inferential-role theories, it should be fairly clear by now how my argument can be extended to apply to these theories.

In addition to establishing that individualistic theories are incompatible with publicity, one of my secondary goals in this argument was to give a picture of how concepts are distributed among thinkers on an individualistic theory, given that they aren't widely shared. On an individualistic theory, the best we can hope for is that instead of many thinkers sharing the same concept, many thinkers each possess a concept that is (in some sense that must be spelled out) very similar to concepts other thinkers possess.

Of course, this highlights an important question that I have not yet taken up at length: What are the theoretical costs of giving up a picture on which many people share many concepts in favor of a picture on which many people possess many similar concepts? This is a question that I must address, and it will be the main focus of my next chapter.

My other secondary goal in this argument was to contrast individualistic theories of concepts with social theories of concepts. Whereas on individualistic theories a concept has features like its semantic value prior to having the feature of being possessed by the thinkers who share it, on a social theory a concept's semantic value and other features depend on who shares it. It should be fairly clear that if a coherent social theory of concepts can be articulated, that theory won't have the same problem with publicity that individualistic theories have.

However, at this point it is probably far from clear that a coherent social theory of concepts can be articulated, and it is also probably far from clear what costs are associated with a social theory of concepts. This is why in the last chapter of this work, once the costs associated with individualism are finally clear, I will sketch a possible version of a social theory of concepts and discuss some potential problems that theory might seem to have. At that point, I will hope to show that most of the apparent costs associated with a social theory actually turn out to be benefits, at least if the theory is meant to do justice to something like our ordinary picture of ourselves as social, thinking creatures.

Chapter 4

Sameness and Similarity of Concepts

My main aim with this project is to motivate a social theory of concepts, where social theories are contrasted with individualistic theories. This chapter will conclude my three-chapter argument that individualistic theories of concepts have a problem with publicity by preemptively responding to a certain objection that an individualistic theorist might make to the argument I made in the last two chapters.

Individualism is incompatible with publicity

To quickly recap the story so far: I began this project with an interest in distinguishing cases of true communication, agreement, or disagreement from cases of merely apparent communication, agreement, or disagreement—from cases in which two or more people are merely 'talking past' one another. I will return to these cases in more detail below, but I hope they are familiar enough both from philosophy and from everyday life. The overall strategy that I am pursuing for distinguishing these cases has a central role for concepts: different people can genuinely communicate, et cetera, if they share concepts, whereas cases of talking past one another are typically cases in which the different people do not share concepts.

One obvious way to determine whether two people share concepts or not is to compare the ways they think involving their respective concepts, giving the verdict that they do share concepts only if, say, there are no important differences in the inferences they are inclined to make with the concepts, or there are no important differences in the objects in response to which they are inclined to use the concepts. This, broadly, is the strategy that comes out of individualistic theories of concepts: examine each thinker in turn, as an individual, to determine which concept she possesses, and then simply ask whether the different thinkers possess the same concept or different concepts.

But this strategy will not be useful if it seldom gives the verdict that two different thinkers do indeed share a concept. That is, if a theory of concepts is incompatible with concepts being widely shared, it will not be of much help in distinguishing genuine communication, agreement, or disagreement from merely apparent communication, agreement, or disagree-

ment. Thus, the publicity of concepts is an important desideratum for at least some theories of concepts.

I briefly offered the above line of reasoning in my first chapter. Then in chapters 2 and 3, I offered a more thorough argument that individualistic theories of concepts are incompatible with the publicity of concepts, because on an individualistic theory the best picture of the actual world we can hope for is one on which many people have many concepts that are merely very similar (at least intuitively) to concepts other people have. This falls short of publicity, which as I have presented it at a minimum involves many people actually sharing many concepts.

Similarity will not offer a solution

My argument in the preceding chapters is thus left open to the objection that many people having many similar concepts is enough for publicity, or at least is enough for the theoretical benefits publicity is supposed to have. So in this chapter I will revisit the question of why and how publicity is a desideratum and then proceed to argue that at least for some important purposes, we should prefer a theory on which concepts are truly public—that is, a theory on which many people share many concepts, where sharing concepts is understood in terms of having the same concepts—over a theory on which many people merely have many concepts that are very similar to concepts other people have.

In the first section of this chapter, I will discuss different kinds of theories of concepts and the different goals that come with these different kinds of theories. The upshot of this discussion will be that there is an important kind of theory—theories that are meant to respect our ordinary picture of ourselves as social, thinking creatures—for which it is an important desideratum that concepts be widely shared in some meaningful sense. However, this discussion will leave open exactly what sense of "sharing" is required. That is, it will leave open whether sharing a concept can be understood as having similar concepts or whether it must be understood as having the same concept.

In the second section, I will compare these two ways of understanding sharing. I will argue that the desideratum of widespread sharing will not be met if, on a theory, it turns out that many people merely have many similar concepts, whereas the desideratum will be met if it turns out that many people truly share many concepts. That is, I will argue that understanding sharing a concept as having similar concepts will not do—where understanding sharing a concept as having the same concept would do—for the purpose of giving a theory that respects our ordinary picture of ourselves as thinking social creatures.

Finally, in the third section I will respond to some objections, both potential and extant in the literature, to understanding concept sharing in terms of sameness rather than similarity.

4.1 Widespread sharing as a desideratum

In the introductory chapter of this work, I gave some idea of why I take publicity to be an important desideratum for a theory of concepts. In this section, I will revisit that discussion, paying particular attention to the various goals one might have in giving a theory of concepts: Even if allowing us to hold on to something like our ordinary picture of ourselves as social, thinking creatures requires some kind of widespread concept sharing, when and why is it important that a theory allows us to hold on to something like that ordinary picture?

4.1.1 Compatibility with our ordinary picture is sometimes important

First, I want to talk about kinds of theory for which respecting our ordinary picture of ourselves is important. Along the way, I will also talk about at least some kinds of theories for which this kind of respect is not important. In the next subsection, I will take up the connection between our ordinary picture of ourselves and concept sharing.

4.1.1.1 Kinds of theories of concepts

There are many different goals one might have in giving a theory of concepts, and the different goals one might have correspond to different kinds of theories one ought to give. For example, if one's main goal is to predict and explain—in a purely descriptive way, rather than in any normative way—the categorization behavior of individuals, one ought to give a theory that is importantly different from the theory one would give if one's main goal were to explain the difference between correctly and incorrectly applying categories.

To get a bit more concrete, suppose that theorist A's main goal is to give a theory that can account for empirical facts related to the ways individuals go about categorizing dogs and non-dogs. These include the fact that many people are a bit quicker to recognize German Shepherds than Chihuahuas as dogs; the fact that when asked to list typical features of dogs, many people will list more features that German Shepherds have than features that Chihuahuas have; and many other facts like these. With a goal like this, it would be entirely appropriate for A to give a theory of concepts along the lines of a prototype or exemplar theory of concepts—a kind of theory that has come to prominence in psychology and cognitive science in the last several decades. Suppose A does give such a theory.

On the other hand, suppose that theorist B's main goal is to give a theory that can distinguish between individuals who are correctly categorizing dogs and non-dogs and individuals who are incorrectly doing so. This will involve specifying what it takes for something really to be a dog—giving some kind of theory of *doghood*—and also specifying what it takes for someone to be wrong about whether something is a dog. B might try to explain someone's being wrong about whether something is a dog in terms of that person's incorrectly applying the concept DOG to a particular object. So with a goal like this, it would be entirely appropriate for B to give a theory of concepts along the lines of Fodor or Peacocke's theory

from the last chapter—a theory that specifies possession conditions and a semantic value for the concept DOG. Suppose B does give such a theory.

Now, there is an important sense in which A's theory is a theory of concepts, and there is an important sense in which B's theory is a theory of concepts. But there is also an important distinction between the way that A's theory is a theory of concepts and the way that B's theory is a theory of concepts. Even though the theories might both use a phrase like "the concept DOG" and use very different descriptions with it, A and B do not have to disagree with one another in giving their theories. Upon hearing about B's theory, A could reasonably say, "That's very interesting, but I'm giving a theory of how people actually categorize dogs, not of how they ought to categorize them." And upon hearing about A's theory, B could reasonably say, "That's very interesting, but I'm giving a theory of what it is to be able to think about dogs and be right or wrong, not of the subpersonal machinery that is involved in categorizing dogs and non-dogs."

The two theories are meant to meet different goals, and these goals come with different desiderata that the theories must meet in order to satisfactorily meet these goals. For example, if there are significant differences among individuals in their categorization behavior with respect to dogs—if some individuals are quicker to recognize German Shepherds than Chihuahuas as dogs, but other individuals are quicker to recognize Chihuahuas than German Shepherds as dogs, say—then B's theory should allow if not account for these differences, but they might not be relevant to A's theory.

On the other hand, suppose that many people are able to make judgments about dogs—either correctly or incorrectly—in spite of many minor differences among these people in their relations to dogs. In this case, A's theory should provide possession conditions for DOG that are coarse-grained enough apply to many people in spite of these minor differences among them. However, this might not be relevant to B's theory, since whether a judgment is about dogs or not might not be a distinction B is concerned with.

4.1.1.2 The importance of our ordinary picture

So depending on the goals one has in developing a theory, one can reasonably end up with a theory of one kind or a theory of a very different kind, with very different desiderata, even though both theories can reasonably be called theories 'of concepts.' I have been claiming that for at least some theories of concepts, publicity is an important desideratum, because publicity of concepts is required for compatibility with our ordinary picture of ourselves as social, thinking creatures. What, then, are the kinds of theories—what are the theoretical goals—for which compatibility with our ordinary picture is important?

In this subsubsection, I will answer this question without focusing on theories of concepts in particular. This is because concepts themselves are typically theoretical entities, and theories of concepts are typically only interesting as parts of larger theories of ourselves—and typically only successful insofar as they help larger theories of ourselves to achieve some theoretical goal or other. So for now, I will focus on the goals we might have in giving a

theory of ourselves more broadly, and in the next subsection, I will take up the question of how theories of concepts are relevant to certain of these goals.

It seems to me that there are many kinds of philosophical theories of ourselves for which compatibility with our ordinary picture is important to one degree or another. To take one important kind, there are theories that are intended as mere systematizations of our ordinary picture of ourselves, requiring revision of the picture only to avoid the most obvious inconsistencies. We might call theories of this kind 'folk psychological' theories.

Another kind that will serve as an example of a limiting case of theories that must be at least somewhat compatible with our ordinary picture to be entirely successful, there are theories that are meant to help us predict and explain actual human behavior at least somewhat; to only posit entities that exist in our minds or in the world in some real sense; and to be as consistent, simple, and productive as possible while meeting the other theoretical goals we set—but to do all of these things while only using terms and making claims that we humans can grasp and can see as relevant to the parts that make up our ordinary picture of ourselves. We might call theories of this kind 'intelligible' theories of ourselves.

The upshot of the last goal I mentioned when describing intelligible theories is that if a theory involves such radical revision of our ordinary picture of ourselves that we cannot, without great effort, actually believe the theory or relate it to our beliefs and practices involved in understanding ourselves and other people in our everyday life, then it does not meet the goal, and so it is not an entirely successful intelligible theory.

For example, suppose we develop a theory of ourselves that has the implication that persons never actually persist through time—that the requirements of numerical identity are never met by a person at one time and a person at another time. Then we have developed a theory that we cannot actually believe in our everyday life, because adopting it would require us to give up ways of talking about and acting towards ourselves and others that we are incapable of giving up. We simply cannot make sense of persons as merely instantaneous entities. In this way, theories of ourselves can turn out to be too radical to count as entirely successful intelligible theories because they are incompatible with terms or claims central to our ordinary picture of ourselves.

Or suppose we develop a theory that in principle does a very good job of predicting and explaining individual behavior but that only makes these predictions by using machinery that is far removed from the ways we ordinarily predict and explain behavior—say, a theory of human behavior which eliminates any important role for terms like "belief" and "desire" and in which all the theoretical machinery works at the level of interactions between neurons. Then we have developed a theory that may well be graspable but that is not clearly relevant to the ways we ordinarily relate to ourselves and others—there is an important sense of "ourselves" in which that is not clearly a theory of ourselves. This is because when we are interested in predicting and explaining behavior in everyday life, we are typically more interested in the actions of persons, entities who importantly have beliefs and desires, than we are in the mere motions of bodies. And although it is fairly clear how interactions between neurons relates to the motions of bodies, it is much less clear how that relates to the actions of persons. In this way, theories of ourselves can turn out to be too radical to count as

entirely successful intelligible theories because they do not use terms or machinery central to our ordinary picture of ourselves.

This is, of course, not to say that there is no place for theories like either of these. But if we are developing a folk psychological theory, or a intelligible theory, or any other kind of theory for which compatibility with our ordinary picture is important, then we will not entirely succeed if we end up with a theory that is so far removed from the terms, claims, and machinery of our ordinary picture of ourselves that we cannot grasp the theory, believe the theory's claims, or relate the theory to beliefs and practices of everyday life. And if we are comparing that theory to another theory that does allow us to do those things, we have some reason—though a defeasible reason—to prefer that other theory, given our goals.

It seems to me that many philosophers—indeed, most I have discussed so far in this work—explicitly or implicitly have goals that put their theories somewhere in the neighborhood of folk psychological or intelligible theories of ourselves.[1] So I hope that it is not very controversial that at least some kind of minimal compatibility with our ordinary picture of ourselves is important to some theories, and I will not devote any more space to discussing the matter at this abstract level.

4.1.2 Compatibility with our ordinary picture requires widespread sharing

So there are theories of concepts of different kinds, making up parts of theories of ourselves with different goals. And sometimes our goals in giving a theory of ourselves require compatibility with our ordinary picture of ourselves. Why should it follow that widespread sharing is a desideratum for some theories of concepts?[2] How does widespread sharing of concepts relate to our ordinary picture of ourselves?

There are of course many parts of our ordinary picture of ourselves, but the part that is relevant to this question is our ordinary picture of ourselves as social, thinking creatures.

[1]For example, Peacocke begins *A Study of Concepts* by suggesting that "[a]nyone who has reflected philosophically on first-person thought, the exercise of a perceptual concept, our grasp of some psychological notion, or our use of an arithmetic concept ... has reason to be concerned with the philosophical study of concepts" (1995, 1), which gives the strong impression that he intends the theory he develops to be clearly relevant to many people's interests in these fairly ordinary phenomena. And although some of the assumptions and goals Fodor lays out at the beginning of *Concepts* (1998) certainly do not seem to be motivated by a desire to respect our ordinary picture of ourselves, some do—and in any case, one of the goals that Fodor lays out quite explicitly is that a theory of concepts must have concepts being widely shared, where that sharing is not a matter of similarity.

[2]Of course, widespread sharing of concepts might be a desideratum of a theory even if that theory does not have respect for our ordinary picture of ourselves among its goals. For example, some theories in disciplines like sociology or history—disciplines in which humans are treated as importantly social, importantly thinking creatures—might enjoy great benefits in simplicity if they can rely on a framework or background that features widespread sharing of concepts. But I am not interested in these theories, both because I am focusing on philosophical theories in this work and because I am focusing on the reasons I myself am interested in widespread sharing as a desideratum, which have more to do with respect for our ordinary picture of ourselves than with theoretical simplicity.

And the answer to this question comes in two parts: first, we need to bring concepts into a theory of ourselves if we want to do justice to the idea that we are thinking creatures; and second, we need those concepts to widely be shared if we want to do justice to the idea that both we and our thinking are—at least sometimes—importantly social.

I will not seriously try to give the first part of this answer. That is, I will not try to give a serious answer to the question of why we would want to give a theory of concepts at all. That question simply falls outside of the scope of this project, since I am just assuming that the best way to explain human thought—including, say, its systematicity and productivity—involves giving a theory in which concepts feature prominently. And this claim does not distinguish me from other philosophers I am discussing this work: proponents of individualistic theories of concepts are still proponents of theories of concepts.

The second part of the answer is, though, more relevant to this project. So although in my first chapter I already laid out some of my reasons for thinking that widespread sharing of concepts is important to our understanding of ourselves as social, thinking creatures, I will quickly review that discussion here.

It is worth reviewing the discussion both because its conclusion is relevant to my argument here and because some of its details will be relevant to my argument soon. In the next section of this chapter, I will compare understanding concept sharing in terms of similarity to understanding concept sharing in terms of sameness. Among other things, I will argue that understanding sharing concepts in terms of similarity will not work because of certain aspects of the role concept sharing plays in our ordinary picture of ourselves.

4.1.2.1 Social thought and shared concepts

Very quickly, then, my idea here is just that there are certain phenomena that are central to our picture of ourselves as social, thinking creatures that—assuming that concepts are the building blocks of thought—can only be easily explained by positing shared concepts. These phenomena include communication, agreement and disagreement, and others.

For example, at its most basic communication involves some message being passed from one party to another. To distinguish communication from miscommunication, we can simply ask whether the message the first party sent is the same as the message the second party received. And when it comes to communication between two people, at the personal level—the kind of communication that is central in our picture of ourselves as social, thinking creatures—there are so many different ways of sending and receiving messages that the only way we can hope to characterize messages in general is as thoughts ('thoughts' construed broadly enough to include, say, intentions as well as beliefs). So communication involves two people having the same thought, so it requires shared concepts.

Before I move on, it is perhaps worth noting (lest I be accused of begging the question) that I do not mean to use "same thought" and "shared concepts" here in strict senses that require true sameness. Instead, I mean to use them to refer to relations between thinkers distinguish true, successful communication from miscommunication, whatever those relations

might amount to.[3] The idea is simply that communication, as do other phenomena, involves this important point of contact between the minds of the people doing it. In the next section, of course, I will argue that this point of contact must amount to sameness, rather than similarity, of concepts. But I do not take that to be settled here.

Moving on from communication, then: Agreement involves two parties assenting to something, and disagreement involves one party assenting to something from which another party dissents. We can distinguish true agreement or disagreement from cases of merely apparent agreement or disagreement in which the first party assents to something that is not the same as what the second party assents to or dissents from. And there is an important sense of "agree" and "disagree" in which two people can agree or disagree not in virtue of what they say but in virtue of what they believe, and believing the same thing requires having the same thought. So agreement and disagreement involve having the same thought, so they require shared concepts.

As I described in the first chapter of this work, I am primarily interested in these phenomena because of my interest in one of the most common everyday situations in which the question arises of whether two people do or do not share a certain concept. That is the situation in which two people are having a discussion and each is using the same word—or at least they are using words that share a sound or spelling—but it becomes unclear whether they are really talking and thinking about the same thing or instead merely talking past one another. For example, it might become unclear whether two people are both talking about 'concepts' in the same sense or whether one person intends "concept" to apply only to certain mental particulars and the other intends "concept" to apply only to certain abstract objects. Or it might become unclear whether two people are both talking about 'justice' in the same sense or whether one person is using "justice" in a sense that primarily involves retribution and the other is using "justice" in a sense that primarily involves distribution of resources.

In situations like these, one way to make clear whether there is genuine communication, agreement, and disagreement or the two people are merely talking past one another is to make clear which concept each is expressing using the word in question. Within the broad theoretical framework that I and many other theorists working in, this primarily involves making clear whether the differences between the two people are differences in which beliefs they have involving a single, shared concept or differences in which concepts they have. For example, this might involve making clear whether one person believes that concepts are (or are best understood as being) mental particulars, whereas the other believes that concepts are (or are best understood as being) abstract objects, and so the two genuinely disagree; or whether each person is using "concept," as well as the concept they are expressing with "concept," as a more technical term for a certain specific theoretical entity, and so the two are actually talking past one another. Or this might involve making clear whether one person

[3]My use of "same thought" here is thus similar to Stich's (1983) use of "content-identity" in that the mere meaning of the term itself is not meant to settle the similarity/sameness question. As I will discuss below, Stich actually argues that "content-identity" must refer to a kind of similarity, rather than a kind of sameness.

believes that justice is primarily a matter of retribution, whereas the other believes that justice is primarily a matter of distribution of resources, and so the two genuinely disagree; or whether each person is using "justice," as well as the concept they are expressing with "justice," in a more specific way to talk about a certain kind of justice, and so the two are actually talking past one another.

Of course, I do not want to claim that the only way to make clear whether two people are talking past one another in this kind of situation involves making clear whether they share a concept. Similarly, I do not want to claim that there is a singular, canonical understanding of each of these phenomena—a single sense of "communicate," "agree," and "disagree"—and that that involves sharing concepts. For example, there is an important and distinct sense of "disagree" on which disagreement does not necessarily require one party assenting to something from which the other dissents, but rather requires each party assenting to something such that it is impossible that both assented-to things are true—in other words, there is an important sense of "disagree" on which both parties disagree just in case they cannot both be right.[4] This sense of "disagree" does not obviously require shared concepts. However, as I discussed in the first chapter of this work, as far as my arguments are concerned, this other sense of "disagree" is not importantly different from the sense in which I am primarily interested.[5]

In any case, there is at least one understanding of communication, agreement, and disagreement that has an important place in our ordinary understanding of ourselves as social, thinking creatures on which these phenomena involve shared concepts. And similarly, one important way to make clear whether two people are talking past one another in this kind of ordinary situation is to answer the question of whether they share a certain concept. So if a theory of concepts is to respect our ordinary understanding and practices, it must allow for concepts to be widely shared in an important sense.

4.2 Problems with understanding sharing in terms of similarity

The upshot of the last section is that, at least for some important kinds of theories of ourselves, compatibility with our ordinary picture of ourselves is important, and our ordinary picture of ourselves involves widespread sharing of concepts. One upshot of the previous two chapters is that individualistic theories of concepts are incompatible with widespread sharing of concepts, where that is understood as many people having many of the same concepts as

[4] I discussed different kinds of disagreement in a bit more detail in the first chapter of this work, and MacFarlane (2007) offers a much more detailed discussion.

[5] To briefly recap that discussion: In the situations in which I am interested, when both parties are in relevantly similar contexts, this other sense of "disagree" will typically require that both parties have concepts with the same semantic values, even if they are not the same concepts. And as should become clear in the next section of this chapter, my complaints about understanding sharing in terms of similarity can be applied at the level of semantic values just as well as they can be applied at the level of concepts.

other people have. Instead, the best picture we can expect on an individualistic theory is one on which many people have many concepts that are in some intuitive sense very similar to concepts that other people have. Of course, this prompts the question of whether we can simply maintain compatibility with our ordinary picture by understanding sharing concepts as having merely similar concepts, rather than as having the same concept.

I want to address this question by discussing two kinds of problems with understanding sharing of concepts in terms of similarity instead of sameness. The first is that as it shows up in our ordinary picture of ourselves, sharing of concepts really is a matter of sameness, rather than similarity, so if we somehow manage to characterize sharing in terms of similarity we will end up giving unintuitive verdicts—verdicts at odds with our ordinary picture—in at least a couple important kinds of cases. The second is that even setting our ordinary picture aside, for certain kinds of theories of concepts there are some serious barriers to using similarity instead of sameness to develop a clear and coherent account of widespread sharing of concepts, because there is no clear way to understand similarity of concepts that does not require us to already have an account on which concepts are widely shared.

4.2.1 Understanding sharing in terms of similarity is at odds with our ordinary picture

It might seem intuitively obvious that our ordinary picture of ourselves as social, thinking creatures involves widespread sharing of concepts[6] and that that sharing is a matter of different people having the same concept, rather than of different people having merely similar concepts. After all, our ordinary picture does seem to involve sharing, and sharing is typically a matter of sameness—of numerical identity, somewhere down the line—in other cases. For example, if two children share a toy or a parent, then there is a single toy or a single parent that each child has. If two people share a name, then there is a single spelling (as "Steven" and "Steven" have), or there is a single pronunciation (as "Steven" and "Stephen" have), or there is a single etymological ancestor (as "Steven" and "Esteban" have), or there is some other single important feature that each person's name has.

At least, it seems intuitively obvious to me that our ordinary picture requires sharing that is a matter of sameness. But in case this impression is not universal—either because I am wrong about just what our ordinary picture includes or because I am not intuiting correctly—I now want to turn the discussion to particular cases and particular theoretical implications. In the two following subsubsections, I will discuss two kinds of problem cases for understanding sharing in terms of similarity. These are cases in which a theory on which sharing is understood in terms of similarity (the kind of similarity that is attainable on an individualistic theory of concepts) will have to give certain verdicts, whereas a theory on

[6] "Concepts" might not be quite the right word here, since the notion of a concept I am using throughout this work is an importantly theoretical notion, and so we wouldn't expect it to show up in what I have been calling 'our ordinary picture' of ourselves, which is supposed to be an importantly pretheoretical picture. But we might expect that that picture has an important role for a pretheoretical analog of concepts, and if it does then that is what I am talking about here.

which sharing is understood in other terms will be able to give different verdicts. In these cases, only understanding sharing in terms other than similarity will allow us to give verdicts that are compatible with our ordinary picture of ourselves.

4.2.1.1 More similar but less shared

The prototype of this first kind of case involves three thinkers, all of whom possess at least somewhat similar concepts: thinker 1 possesses concept C_1, thinker 2 possesses concept C_2, and thinker 3 possesses concept C_3, with concepts C_1, C_2, and C_3 being fairly similar to one another if not the same as one another—though there are, of course, certain differences among the thinkers. The question to ask about the case is the question of whether 1 shares a concept with 2, 3, both, or neither. The intuitive answer to this question will be that 1 shares a concept with 2, because the only interesting difference between 1 and 2 is that 2 has a certain false belief that 1 does not have; whereas 1 does not share the concept in question with 3, because the difference between 1 and 3 amounts to a difference in concepts, rather than a mere difference in beliefs.

However, this way of describing the case begs the question against the individualistic theorist of concepts, because this description will not be available to an individualist. This is because given what an individualist takes concepts to be, there is no way in which C_1 is more similar as a concept to C_2 than it is to C_3. Thus even if sharing is understood in terms of similarity, an individualist will only be able to describe the case as one in which 1 shares C_1 with neither 2 nor 3, with 3 alone, or with both other thinkers.

Thinkers 1 and 2 The details of the case, then: Thinkers 1 and 2 are members of the same linguistic community, and they respectively express their concepts C_1 and C_2 using the word "cat." 1 is an absolute master of using "cat" and of differentially responding to *cathood* in thought using C_1. Intuitively, C_1 is CAT. And on any reasonable theory of concepts, if any of the three concepts in this case is CAT, C_1 is (though that is not to rule out that more than one of the concepts in this case might be CAT).

On the other hand, 2 does not use "cat" in exactly the same way, and 2 does not have thoughts that perfectly track *cathood*. 2 tends to say things like "All cats have fur. Part of being a healthy adult cat is growing fur."and to have analogous C_2-involving thoughts. 2's use of C_2 does not quite track *cathood*, but it does track the property *being a naturally furry cat*. Intuitively, we would describe 2 as being perfectly capable of thinking about cats as such but as having the false belief that all cats have fur (indeed, that they necessarily have fur, at least in some sense). Intuitively, in possessing C_2, 2 shares a concept with 1.

Of course, the case so far should be familiar from the last chapter, as should the verdict that an individualistic theory must give. 2 makes different concept-individuating inferences with C_2 than 1 makes with C_1, and 2 tracks a different property with C_2 than 1 tracks with C_1, so on an inferential-role or causal individualistic theory of concepts, C_2 will not be the same as C_1, and 2 will not have the false belief about cats mentioned above.[7]

[7]I will not review in detail my discussion from the last chapter of attempts by individualistic theorists

However, an individualistic theorist might try to give something like the intuitive verdict concerning 1 and 2 by proposing that we understand sharing of concepts in terms of similarity. If the individualistic theorist could propose a coherent similarity metric and a method for determining how much similarity is required for sharing a concept, it may well turn out that C_1 and C_2 are similar enough that 1 and 2 are, on that proposal, counted as sharing a concept. In the next subsection, I will question whether any coherent and useful similarity metrics are available to an individualistic theorist, and one could question whether C_1 and C_2 would really be similar enough to reasonably count as being shared, but for now I will simply assume that this proposal could be made to work for 1 and 2.

Thinker 3 Consider, though, thinker 3, who does very specialized research in genetics. 3's entire career is concerned with a very very common gene with very subtle effects that shows up in species throughout the animal kingdom. More specifically, 3's career is concerned with the effect on an animal of possessing this gene (a very very common but very subtle phenomenon). And 3 is so devoted to this research that it is as if all of 3's time is spent thinking and talking of animals only in terms of those who possess this gene, including when it comes to cats. Except when it comes to the gene, 3's use of "cat" and C_3 are exactly like 1's use of "cat" and C_1. But 3 uses "cat," expressing C_3, in such a way that 3 tends to say things like "All cats have the gene. Part of being a cat is having the gene." and to have analogous C_3-involving thoughts. 3's use of C_3 does not quite track *cathood*, but it does track the property *being a cat with the gene*.

So far, 3 should sound very like 2: both use "cat" in the same way that 1 does except when it comes to a certain rare subclass of cats (cats without the gene in the case of 3, and hairless cats in the case of 2). However, there are a few important differences between them. Unlike 2, 3 recognizes that most other people use "cat" slightly differently and does not think that people who use "cat" differently are using the word incorrectly. In fact, 3 is more deferential in using "cat" than 2 is—except when the gene is at issue—because 3 is more humble than 2 generally. And when 3 hears other people use "cat," 3 often has to remember to understand their utterances in part by using a complex concept like ANIMAL THAT IS EITHER A C_3 OR OF THE SAME SPECIES AS A C_3 BUT LACKING THE GENE, whereas 2 almost always understands "cat"-involving utterances just by using C_2. And lastly, another important difference between 3's situation and 2's situation with respect

like Fodor and Peacocke to solve this problem, showing how their theory really does give the intuitive verdict in a case like this or amending their theory so that it gives something like the intuitive verdict in a case like this. However, allow me to add to the case that (at least in the situations in which 2 ends up using C_2) 2 is no more inclined to defer to thinkers like 1 than (in the situations in which 1 ends up using C_1) 1 is to defer to thinkers like 2. This blocks Fodor's attempted solution to the problem, which relies on pervasive and asymmetric deference, from applying to this case. Further allow me to point out that the intuitive verdict is that, at least in some important sense, 1 and 2 are both thinking about cats as such when they use C_1 and C_2, but they are not both thinking about naturally furry cats as such when they use the concepts—intuitively 1 is using C_1 correctly, but 2 is not using C_2 correctly. This blocks Peacocke's attempted solution to the problem, which fails to distinguish between 1 being right and 2 being right, given that they express their concepts using the same word.

to cats is that possessing the gene is much more common in cats than being naturally furry, and the effects of possessing the gene are much more subtle in cats than being naturally furry.

Intuitively, C_3 is not the same concept as C_1, and in possessing C_3 3 does not share a concept with 1. Instead, 3 uses "cat" as a somewhat technical term in a somewhat specialized dialect, and 3 shares C_3 only with others who speak the same dialect. But if concepts as similar as C_1 and C_2 count as shared just in virtue of their similarity, C_1 and C_3 will also count as shared unless there is some respect in which C_1 and C_3 are less similar than C_1 and C_2 that is relevant to the individualistic theorist's proposal under discussion here.

Similarity This, of course, depends on just how similarity is proposed to be relevant to concept sharing. Ideally, an individualistic theorist could propose some similarity metric that allows for easy comparison of the similarity of any given pair of concepts against any other pair of concepts or against some standard of similarity such that any pair of concepts counts as shared of it meets the standard. Of course, such a straightforward metric would likely be a significant oversimplification, since the complexity of concepts might not allow for comparison of similarity along a single dimension, and standards of similarity might vary across contexts. However, I am not interested in these problems—I am simply interested in whether there is any way an individualistic theorist could compare the similarities of the three concepts in this example to ultimately use similarity to give the intuitive verdict about concept sharing in this case.

So I will now discuss the options that are available to an individualistic theorist when it comes to similarity metrics, no matter how complex those metrics might be. This discussion, though, will be brief, since I want to present these options as charitably as I can, and as I will show in the next subsection, upon closer inspection it becomes clear that some of these options cannot work as intended.

The most natural proposal for a similarity metric would be that two concepts are similar insofar as they have similar individuating properties: properties in virtue of which they are the concepts they are. So on a causal individualistic theory of concepts, concepts might count as similar insofar as the properties to which they have an individuating causal relation are similar. On a proposal like this, C_1 and C_2 will count as very similar because the properties *being a cat* and *being a naturally furry cat* are very similar, perhaps because very few objects have one property but not the other or because they are complex properties that share most of their parts and structure.

Similarly, on an inferential-role individualistic theory of concepts, concepts might count as similar insofar as they have similar individuating roles in the thoughts of individuals who possess them. On a proposal like this, C_1 and C_2 will count as very similar because their roles share many inferences and only differ when it comes to inferences like THAT IS A C, THEREFORE THAT NATURALLY HAS FUR.

At least at first glance, then, it might seem feasible to count C_1 and C_2 as very similar. But how do C_1 and C_3 compare on these similarity metrics? As I have described C_3, it is

hard to see how it could be counted as less similar to C_1 than C_2 is: The property to which C_3 is individuatingly causally related is at least as similar to C_1's property as C_2's property is, since there are even fewer cats without the gene than there are naturally furless cats and since (I feel safe in stipulating) *being a cat*, considered as a complex property, shares more parts and structure with *being a cat with the gene* than it does with *being a naturally furry cat*. And because the effects of the gene are so subtle, having the gene licenses fewer inferences than having fur, so if anything the inferential role of C_1 shares more with C_3 than it does with C_2. So it looks like the individualistic theorist still cannot give the intuitive verdict about this case, even if we accept the proposal to understand sharing concepts in terms of similarity.

Limitations of individualism Of course, I have not given an exhaustive taxonomy of similarity metrics for concepts available to individualistic theorists. There are other potentially individuating features of concepts that could play into similarity on some theories, such as a concept's sense or mode of presentation. There are also many non-individuating features of concepts that could play into similarity. But it should be clear upon reflection that the features of a concept that could be individuating on a reasonable individualistic theory will all work more or less like the individuating features already discussed, at least when it comes to the present case. And it would be hard to motivate an account on which concepts are shared in virtue of being similar in non-individuating features: Why should it be important whether concepts are similar in ways that are not relevant to which concepts they are?

The only features that could help a theorist of concepts give the intuitive verdict in this case—that 1 shares C_1 with 2 but not with 3—are the features that distinguish 3 from 1 and 2. What makes 3 different, intuitively, is that 3 recognizes that 3's use of "cat" and of C_3 is somewhat unique and that this uniqueness is appropriate, because 3 does not intend to share C_3 with thinkers like 1 in virtue of their possessing C_1. So if a theory could somehow take these social features of C_3 into account, it would stand some chance of giving the intuitive verdict.

But, of course, taking social features like these into account is antithetical to the individualism of individualistic theories. And since an individualistic theory will not count social features as individuating features of a concept, on an individualistic theory there is a very important sense in which they are not even features of the concept—they will instead be counted merely as features of (perhaps some) thinkers who possess the concept. This is because if social features are not individuating, it is possible for a thinker to possess the same concept without having those same social features.

For example, consider a thinker who possesses a concept that is causally connected to the same property as C_3 and that has the same inferential role as C_3 not because the thinker has a unique concept that they recognize as such but because the thinker has heard about the gene and come to the mistaken belief that having the gene is part of being the cat (because, say, the thinker heard that animals cannot survive without the gene). On an individualistic

theory, then, this thinker will likely count as possessing C_3.

So if an individualistic theory counts a thinker as possessing C_3 whether or not that thinker intends to share the concept with thinkers like 1 in virtue of their possessing C_1, then the theory cannot count that as a feature of C_3 and so somehow count C_3 as less similar to C_1 than C_2 is. So however individualistic theorists propose to understand similarity of concepts, they cannot give the intuitive verdict in this case by understanding sharing concepts in terms of similarity of concepts.

Broader implications Before I leave this case, I want to make one final point. It is likely that some readers will not share my intuitions about this particular case. And even those readers who do share my intuitions will likely be reluctant to put a great deal of argumentative weight on these intuitions that are, after all, about a single contrived case that has been described without much detail. But whether a theory of concepts gives what I have been calling 'the intuitive verdict' about this single case or not, it is surely important that the theory allows cases that are like this one to occur and to occur widely.

There are two important features of this case that theories of concepts ought to allow. The first feature is that in this case, a thinker (like 2) can count as possessing and sharing a concept even though that thinker makes some mistakes with it. If a theory requires full mastery of a concept for being able to use that concept—for example, if the theory does not allow a thinker to use CAT if that thinker is not immediately open to the possibility of hairless cats—then given our human limitations, on that theory we will rarely be able to use any concept of any significant complexity. Being able to use concepts without mastering them is a very important part of any realistic picture of human thought.

The second important feature of this case is that a thinker (like 3) can count as not using a concept even though the thinker possesses a very similar concept. If a theory counts any similar concept as shared—for example, if the theory counts anyone thinking about cats possessing a very common gene as sharing thoughts with someone thinking about all cats—then given the subtle distinctions we often have to make in scientific and philosophical theorizing, that theory will often fail to distinguish between two theorists disagreeing about a certain subject area and two theorists proposing subtly different ways to understand that subject area. Being able to use idiosyncratic or technical concepts that are similar to other concepts but still different from them is also a very important part of any realistic picture of human thought.

In the last two chapters, I argued that individualistic theories of concepts cannot allow cases to exhibit the first of those two features, since these theories in effect equate concept possession with concept mastery. Now, as I am discussing in this chapter, an individualistic theorist might try to avoid some of the problems that follow from equating possession with mastery by proposing to understand concept sharing in terms of concept similarity. But the upshot of this subsubsection is that such a theorist will thereby disallow some important cases from exhibiting the second of those two features. On an individualistic theory of concepts, similarity-dependent concept sharing will either be too weak to allow enough sharing in cases

where thinkers fall short of concept mastery or too strong to allow enough nonsharing in cases where thinkers use concepts in a deliberately idiosyncratic or technical way.

4.2.1.2 Similarity and transitivity

In the second kind of case I want to discuss, the question of whether concepts are shared again comes apart from the question of whether concepts are similar, as it did in the case above. Here, though, the distinction is more structural. The problem comes down to this: sharing, as it is naturally understood in terms of sameness, is a transitive relation, but similarity is not transitive.

I already discussed transitivity in the first chapter of this work, so I will not give a very detailed discussion here. And I will also discuss transitivity in the next section when I discuss objections to understanding sharing concepts in terms of sameness. In this subsubsection, then, I will just try to quickly highlight the difference that transitivity of sharing concepts makes to our picture of ourselves as social, thinking creatures.

First, then: similarity is not transitive. I hope that this is not controversial, but to help make the point clear, again consider three thinkers, each of whom possesses a concept of interest. Thinker 4 possesses concept C_4, 5 possesses C_5, and 6 possesses C_6. C_4 is similar to C_5, and C_5 is similar to C_6. Does it follow that C_4 is similar to C_6?

No, it does not follow. The features in virtue of which C_4 is similar to C_5 might be completely distinct from the features in virtue of which C_5 is similar to C_6—C_4 might have no interesting features in common with C_6. To compare to similarity in another area, if one book is similar to a second book because they are written in the same style, and the second book is similar to a third book because they concern the same subject matter, it clearly does not follow that the first book is similar to the third book. And since concepts, just like books, can differ from one another in many different ways, they also exhibit this kind of non-transitivity.

Or, returning to concepts, if C_4 is similar to C_5 and C_5 to C_6 because there is some single feature that C_4 and C_5 have to similar degrees and that C_5 and C_6 have to similar degrees, it might be that the degree to which C_4 has the feature is significantly dissimilar from the degree to which C_6 has that feature. To compare to similarity in another area again, if one book is similar in length to a second, and the second to a third, it does not follow that the first is similar in length to the third—for example, if 400 pages is similar as a length to 500 pages, and 500 pages to 600 pages, it does not follow that 400 pages is similar as a length to 600 pages. And since concepts, again like books, can have features that come in degrees, they also exhibit this kind of non-transitivity.

So similarity of concepts is not transitive. Sharing a concept, though, is transitive, and importantly so, on our ordinary picture. When a teacher successfully imparts a concept to a student, and that student grows up to become a teacher and successfully imparts the concept he or she acquired to a new student, and so on, intuitively there is a single concept passed down the teacher-student chain so that thanks to the chain a student reading the work of the long-dead original teacher can count on being able to think about the same

things the original teacher was thinking about. And when a single speaker expresses a belief in front of a wide audience and many members of the audience agree with the speaker's belief, thinking with concepts they share with the speaker, they can count on agreeing with one another simply in virtue of agreeing with the speaker. These phenomena are simply part of our ordinary picture of ourselves, and that they work the way they do requires that concept sharing be a transitive relation.

This is not to say that there can be no coherent picture of a phenomenon like communication, or agreement and disagreement, founded on concept sharing understood in terms of similarity. But to adopt such a picture would be to significantly revise our ordinary picture of ourselves as social, thinking creatures, because on that picture we could no longer count on the thought at one end of a chain of communication or agreement being the same as—or even similar to—any thought at the other end, as long as that chain involved more than two thinkers.

4.2.2 For inferential-role theories, similarity depends on sameness

In this section, I am discussing problems with understanding sharing concepts in terms of having similar concepts. In the last subsection, I focused on such problems that arise from a mismatch between the picture that understanding gives us and our ordinary picture of ourselves as social, thinking creatures. In this subsection, I will discuss a different kind of problem: problems with understanding similarity of concepts in any coherent, useful way that does not depend on many people already having the same concepts.

If the notion of similarity of concepts is to be useful in an individualistic theory in the present dialectical circumstances, then that theory must be able to account for many people having many concepts that are similar to concepts other people have. But as I have already discussed, individualistic theories of concepts are not able to account for many people having many of the same concepts. So if an individualistic theory is to make use of similarity, it must be able to account for that similarity such that widespread similarity does not depend on widespread sameness.

However, inferential-role theories of concepts face significant difficulties in accounting for similarity of concepts in a way that allows for widespread similarity and does not depend on widespread sameness. This is not to say that other theories can easily account for similarity—it is not entirely obvious to me how, for example, a causal theorist should understand similarity of concepts. But I will focus on this problem as it applies to inferential-role theories because it is more serious than the analogous problem for causal theories and because it has seen more discussion in the literature, particularly in Fodor's writing.[8]

[8]Fodor and LePore (1992) offer an argument, which Fodor later summarizes (1998), that is at its heart very similar to the one I am making in this subsection.

4.2.2.1 Sharing many inferences

As I mentioned in the last subsection, it is most natural to account for similarity of concepts in terms of similarity in their individuating features. On an inferential-role theory, then, it is most natural to account for similarity of concepts in terms of similarity of the individuating parts of their inferential roles. At least at first, this line of thought might seem very promising, because it is fairly clear how an account like this could be applied to individual cases.

For example, if a theorist with this account compares two thinkers, A and B, who have (intuitively only subtly) different concepts they both express using the term "cat," C_A and C_B, it is fairly clear how the theorist could count those concepts as similar. If 99 of the 100 kinds of inferences A makes with C_A that determine which concept C_A is are kinds of inferences that B makes with C_B that determine which concept C_B is—if, say, the individuating inferential roles only differ when it comes to inferences of the form THAT IS A *C*, THEREFORE THAT NATURALLY HAS FUR, which after all should comprise a relatively small part of the individuating inferential role of a concept expressed with "cat"—then the theorist will count C_A and C_B as similar.

However, given that two thinkers having the same concept as one another is less common on an individualistic inferential-role theory than one might have supposed, as I argued in the last two chapters, it is worth taking up the question of how common it will be on such a theory that two thinkers make the same kinds of inferences with their concepts. For example, how common will it be on such a theory for two thinkers to both make inferences of the form THAT IS A *C*, THEREFORE THAT IS A MAMMAL?

Presumably, among other things, making an inference of that form requires having the concept MAMMAL. So making an inference of that form will not be more common than possessing MAMMAL. But given that MAMMAL is the kind of high-level, complex concept that was the focus of my argument in the last two chapters, possessing MAMMAL will turn out to be very rare on an individualistic inferential-role theory of concepts. This is because as such a high-level, complex concept, MAMMAL involves enough subtlety in its individuating inferential role that we should not expect many thinkers to have a concept with precisely (or imprecisely) that individuating inferential role.

Since we should not expect many thinkers to possess MAMMAL on an individualistic inferential-role theory, we should not expect many thinkers to make inferences of the form THAT IS A *C*, THEREFORE THAT IS A MAMMAL. And given the complexity of the other concepts likely involved in the individuating inferential role of any concept expressed using "cat," we should not expect the two given concepts, C_A and C_B, to be involved in many of the same kinds of individuating inferences. So if similarity is being involved in many of the same kinds of individuating inferences, given that C_A and C_B were not specially chosen, we should not expect two thinkers having similar concepts to be widespread on an individualistic inferential-role theory.

4.2.2.2 A proposal and a problem

Of course, there are other ways an individualistic inferential role theorist could propose to understand similarity. One is as being involved in many similar kinds of individuating inferences, rather than as being involved in many of the same kinds of individuating inferences. And, indeed, since we are already considering making use of similarity rather than sameness when it comes to concepts, it would be natural to move to similarity rather than sameness when it comes to forms of inferences.

How, then, might we understand similarity of forms of inference? Upon a little reflection, it should be clear that this is a fairly complex notion. Merely involving similar concepts—or even all the same concepts—is not enough, at least if inferences like THAT IS A *C*, THEREFORE THAT IS NOT BOTH FULL-GROWN AND SMALLER THAN A MOUSE and THAT IS A *C*, THEREFORE THAT IS BOTH NOT FULL-GROWN AND SMALLER THAN A MOUSE involve the same concepts. As those inferences illustrate, at least on some reasonable understandings, there is significantly more to an inference than which concepts it involves.

But whatever inferences involve beyond their concepts, it should be relatively uncontroversial that in many cases, involving dissimilar concepts should rule out inferences as being similar: THAT IS A *C*, THEREFORE THAT IS A MAMMAL is importantly dissimilar from THAT IS A *C*, THEREFORE THAT IS A REPTILE precisely because MAMMAL is importantly dissimilar from REPTILE. And when it comes to the two thinkers A and B, who both express their respective concepts MAMMAL_A and MAMMAL_B using the term "mammal," it should be fairly uncontroversial that inferences of the forms THAT IS A C, THEREFORE THAT IS A MAMMAL_A and THAT IS A *C*, THEREFORE THAT IS A MAMMAL_B will be dissimilar insofar as MAMMAL_A is dissimilar from MAMMAL_B.

So if an individualistic inferential-role theorist is trying to make use of the notion of similarity of forms of inference, hoping to count A and B as possessing similar concepts each expresses using "cat," then that notion could prove useful only if enough of the kinds individuating inference A makes involving C_A are similar enough to kinds of individuating inference B makes involving C_B. And assuming that these similarity comparisons work the way the comparisons in the above paragraph worked, that will require that A has many other concepts that are similar to concepts B has, such as MAMMAL_A.

So C_A will count as similar to C_B only if many related concepts A possesses, like MAMMAL_A, count as similar to related concepts B possesses, like MAMMAL_B. But, then, each of those pairs of related concepts will count as similar only if many more pairs of concepts count as similar, and so on.

Similarity of a single pair of concepts, then, ends up being a very fragile and demanding phenomenon on a theory like this. This is not to say that it is an impossible phenomenon, but it is hardly as dependable as a theorist hoping to found publicity on it might like. Given the interconnectedness of our conceptual webs, if there is some subject matter that A and B think about in significantly different ways—for example, if they think very differently from one another about religion, politics, or some part of the natural world—then there is

a serious danger that the dissimilarity of their concepts relating to that subject matter will infect most of their other high-level concepts on a theory like this.

4.2.2.3 A further problem and a simple example

Additionally, it is not clear that this kind of similarity founded on similarity (founded on similarity ...) will, even when it occurs, be enough to ground any reasonable understanding of sharing concepts. I will illustrate this problem with a toy example.

Suppose that thinker A possesses concept C_A and thinker B possesses C_B, and consider the question of whether C_A and C_B are similar. Each of C_A and C_B are involved in one kind of individuating inference: inferences of the form THAT IS A *C*, THEREFORE THAT BEARS THE RELATION R_C TO THE KIND *D*, where in these inference forms the concepts *C*, R_C, and *D* vary between the two thinkers. C_A is thus related by individuating inference to the concepts R_{CA} and D_A, and C_B is related to the concepts R_{CB} and D_B. Each of the thinkers' *D* concepts is in turn involved in a single similar kind of individuating inference, relating it to R_D and *E* concepts: THAT IS A *D*, THEREFORE THAT BEARS THE RELATION R_D TO THE KIND *E*. And each of the *E* concepts is individuated according to its relations to R_E and *F* concepts, and so on, down to *M* concepts ten 'levels' of inference from the *C* concepts in which we are interested.

On an inferential-role theory, then, the identities of the two thinkers' *C* concepts are determined by the identities of the *M* concepts and all the *R* concepts in each thinker's conceptual web.[9] This is because the identity of C_A is determined by the identities of R_{CA} and D_A, but the identity of D_A is determined by the identities of R_{DA} and E_A, and so on to the identities of R_{LA} and M_A—and the same goes for B's concepts.

And on the kind of theory here under discussion, two concepts related by individuating inferences to other similar concepts—related as the *C* concepts are to the R_C and *D* concepts are in this example—count as similar. So C_A counts as similar to C_B just in case each of A's ten *R* concepts is similar to its counterpart among B's *R* concepts and M_A is similar to M_B.

For this example to achieve its main aim, it requires the further supposition that A's *R* concepts each count as similar to its counterpart among B's *R* concepts. But before I continue further towards that main aim, let me briefly highlight the fragility of similarity of concepts on the kind of theory I am here considering by highlighting the importance of this supposition.

If only one among A's ten *R* concepts is not similar to any of B's concepts, then given the inferential structure already supposed, that will be enough for C_A to count as dissimilar from C_B and from every other concept B possesses. For if, say, R_{HA} is not similar to R_{HB} or to any other concept B possesses, then B will not be able to possess a concept similar to H_A, since making an inference similar to H_A's individuating inference will require possessing

[9] Of course, this is not strictly true, since as I gave the individuating inferences above, they involve other concepts like BEARS. But these are not particularly interesting or important concepts in this dialectical context, and we can simply suppose that the two thinkers share these other concepts. The important question of how C_A relates to C_B, then, does not depend on these other concepts in any interesting way.

a concept similar to R_{HA}—and failing to possess a concept similar to H_A will disqualify B from possessing a concept similar to G_A, which will disqualify B from possessing a concept similar to F_A, and so on up to C_A. So even on this very permissive theory of similarity of concepts, given the interconnectedness of our conceptual webs, it is by no means guaranteed that a given thinker possesses a concept similar to a given concept another thinker possesses.

However, I would like to leave that worry aside and simply assume that each of A's R concepts is similar to its counterpart among B's R concepts and that M_A is similar to M_B. It then follows, on the kind of theory here under discussion, that C_A is similar to C_B. But should this really follow, if similarity is to ground an understanding of sharing concepts?

I do not think this should follow, since it seems to me that it is possible for the differences between the counterpart concepts in two inferential webs to grow the farther along the webs we look, if at each step we are only guaranteed that the inferences in the one web are similar to the inferences in the other web. This is largely because of a phenomenon closely related to the non-transitivity of similarity.

4.2.2.4 A simple model for the simple example

To see this as plainly as possible, it might help to use a mathematical model of concept similarity. Imagine kind concepts like the C through M concepts as points in some conceptual space, and imagine relational concepts like the R concepts as vectors in that space that locate one point relative to another. For the sake of simplicity, imagine that this space is two-dimensional and so that each point in this space can be identified with an ordered pair of numbers, such as "(3,4)." So if M_A is the point (0,0), and R_{LA} is the vector $\langle 20,0 \rangle$, then since L_A is related to M_A by R_{LA} in its one and only type of individuating inference, L_A will be the point (20,0).

Now, this model is clearly much too simple in several respects to be of general use in studying concepts. But realistically complicating the model—by, say, making the space massively multidimensional, treating kind concepts as partial functions from points in this space to values of some kind, and treating at least some relational concepts as partial functions from those functions to others, for a start—would greatly complicate this discussion but would not affect the upshot of this example that I am interested in here. And that upshot is simply that small, even apparently insignificant dissimilarities can accumulate as we move through an individual's conceptual web.

For one last useful oversimplification, assume that two concepts are similar if the distance between their points or the magnitude of the difference between their vectors is less than 5. It is probably clear, then, where this example is going, but I will quickly spell it out.

Let M_A be (0,0) and M_B also be (0,0). Let R_{LA} be $\langle 20,0 \rangle$, but let R_{LB} be $\langle 19,0 \rangle$. Furthermore, let the rest of A's R concepts each be $\langle 1,0 \rangle$ more than its counterpart among B's R concepts, and let the rest of B's R concepts each take any value—it does not matter what. It follows, then, that even though M_A is the same as M_B, and even though each of A's R concepts is very similar to its counterpart among B's R concepts, C_A will be 10 units of distance from C_B, which is well above the usual threshold for similarity.

4.2.2.5 A slightly more complicated model and example

Of course, this particular way that dissimilarity accumulates is in part an artifact of the proposed model and the details of the example. To show that as far as dissimilarity accumulation goes, this model and example are actually quite conservative, allow me to slightly change them. Now imagine relational concepts like the R concepts as more complicated functions from points to points in the same conceptual space, instead of as simple vectors. And to avoid needing some way to distinguish similar from dissimilar pairs of functions, let each of A's R concepts have exactly the same function as its counterpart among B's R concepts. To keep things very simple, let every R concept have the function that, given the point (x, y) as input, returns the point ($2x$, $2y$).

If M_A is (1,0), then, and M_B is (0,1), L_A will be (2,0) and L_B will be (0,2). And L_A will be counted as similar to L_B because their individuating inferences involve only similar concepts: M_A is similar to M_B, and R_{LA} is similar to R_{LB}. Moving on, K_A will be (4,0) and K_B will be (0,4), and they will be counted as similar because the L and R_K concepts are similar. Skipping to the end, C_A will be (512,0) and C_B will be (0,512), and they too will be counted as similar, this time because the D and R_C concepts will be similar.

So with this slightly more complicated model, even if A and B have exactly the same R concepts, a difference in their M concepts of one unit in each dimension can grow to a difference in their C concepts of more than five hundred units in each dimension—again well above the usual threshold for similarity.

4.2.2.6 The upshot of these examples

The upshot, then, is that having similar individuating inferences should not guarantee that two concepts are similar, because even very small dissimilarities among concepts can accumulate as one moves through conceptual webs—and as I argued in the last two chapters, on an individualistic theory of concepts we should expect many very small dissimilarities among the concepts of two different thinkers. So an individualistic inferential-role theory of concepts that uses such a criterion for similarity of concepts and uses similarity of concepts as a criterion for sharing concepts will give undesirable verdicts in many cases.

Faced with this problem, a proponent of such a theory might try to propose some method for distinguishing between the very small dissimilarities that accumulate and those that do not as one moves through conceptual webs, or to propose some method for determining how much dissimilarity accumulates as one moves through conceptual webs. If successful, the theorist could then use this method to distinguish concepts with similar inferential roles that truly are similar from concepts with similar inferential roles that are not truly similar, and so refrain from giving as many undesirable verdicts about sharing.

But it seems to me that there is little prospect for an inferential role theorist to come up with a palatable method for doing either of these things, at least if that theorist wishes to maintain that a concept should be identified by its inferential role. This is because distinguishing between similar roles that make for shared concepts and similar roles that do

not make for shared concepts would greatly lessen the importance of the inferential role of a concept. If a theorist (as an inferential-role theorist) holds that which concept an individual possesses is simply a matter of what role that concept has in the individual's thought and (as an individualist) holds that whether two individuals share a concept is simply a matter of which concept the one possesses and which concept the other possesses, then it would be very strange to propose that concept sharing depends on possessing similar concepts but that possessing similar concepts is not a matter of possessing concepts with similar roles. This is not to bring out any clear logical inconsistency in this proposal, but I think it should be clear that the more weight a theorist puts on the inferential role of concepts, the less attractive this proposal will be.

Very generally, this problem arises for the individualist inferential role theorist because that theorist treats concepts simply as nodes in a network, and it can be very difficult to compare the positions of two nodes in two different networks if there is no way of independently locating nodes other than by their relations to other nodes in their respective networks. If one is tasked with comparing nodes across different networks, one's job will be much easier if there is some way of locating nodes with respect to some third kind of space—for example, if the nodes are concepts that importantly have semantic values relating them to properties or other nonconceptual entities, as they might on a causal theory or anything other than a purely inferential-role theory—or if there are direct relations of some kind between nodes in different networks—for example, if the nodes are concepts that can be related to one another by some kind of interpersonal sharing relation that does not entirely depend on their positions in their respective networks, as on a social theory.

To summarize, then, there are a few ways that an individualistic inferential role theorist might try to understand similarity of concepts, but each has its problems. If the theorist claims that similarity is a matter of involving many of the same individuating inferences, then the requirement will be too strict, because it will be very rare for two individuals to be able to make the same inferences, since that requires having the same concepts. On the other hand, if the theorist claims that similarity is a matter of involving merely similar individuating inferences, then the requirement will be too lax (perhaps as well as too strict), because that claim discounts the possibility of dissimilarity accumulating as one moves through conceptual webs. And lastly, if the theorist claims that similarity is a matter of something else, beyond inferential roles, then that claim will be in tension with the theorist's central claim that a concept's identity is determined by its inferential role.

4.3 Objections to understanding sharing in terms of sameness

Before I end this chapter, I would like to discuss a few potential objections to understanding sharing concepts as having the same concept that are potentially relevant to my argument. These will mainly be objections to my claim that there is an important place for this un-

derstanding in our ordinary picture of ourselves as social, thinking creatures. For the most part, these objections seem to be motivated by a recognition that there is an important place for something like concept similarity in our ordinary picture—a recognition I do not wish to deny. However, the main thrust of these objections seems to be the inference from this recognition to the claim that there is no important place for something like concept sameness in our ordinary picture—a claim I certainly do wish to deny.

Before I discuss these objections, though, I will briefly mention a very important objection I will not be discussing here. Having read this work so far, one could very reasonably object to my argument along the following lines:

Objection: You have argued that individualistic theories of concepts can only account for many people having many concepts that are similar to concepts many other people have, that respecting our ordinary picture of ourselves requires accounting for many people having many of the same concepts as many other people, and that theories of concepts that cannot respect our ordinary picture come with a cost. But I can grant you all of these claims and yet reasonably prefer individualistic to social theories of concepts, because social theories of concepts will come with greater costs.

I will not be discussing this objection in this chapter. Instead, I will address it with the next chapter, where I will describe how the details of a social theory of concepts might go and discuss the costs that come with a social theory. Only once the costs associated with both kinds of theory have been made clear will it be reasonable to compare them.

4.3.1 Similarity is what we have

4.3.1.1 Objection

Particularly if one is suspicious of the distinction between features of a concept and beliefs involving that concept[10], or if one is suspicious of the theoretical benefits of publicity of concepts that I outlined above, one might be inclined to object to my argument along the following lines:

Objection: Given the many subtle differences among people that you made so much of in the last two chapters, why should we expect or demand that concepts possessed by different people ever be the same, rather than merely similar? In everyday life and in philosophical theorizing, one of the lessons we must learn as we come to understand

[10]I am concerned here with a theorist who wishes to claim that *any* belief in which a concept is involved is a feature of that concept. I do not mean to include in this category theorists who hold that merely *some* of the beliefs in which a concept is involved are features of that concept, which is an importantly different claim. For example, a theorist might hold that involvement in the belief TWO IS THE SUCCESSOR OF ONE is an important feature of the concept TWO—that someone who does not have that belief cannot have the concept TWO, say—without holding that involvement in the belief I HAVE TWO CATS is a similarly important feature of the concept that I express using the word "two," even if that concept is involved in that belief.

> how richly varied people are is that no one else thinks quite like we do: for practically any topic of significant complexity, if we look hard enough we will find some difference between the way we think about the topic and the way some other given person thinks about it. From this fact, shouldn't we draw the lesson that different people rarely have the same concepts but that having similar concepts is good enough, since we seem to be doing fine with that?

Now, in a way this is an objection to my argument from the last two chapters, and my response to it is just this chapter. But there is a bit more I can do to respond to this objection, particularly given what seems likely to be motivating this objection—at least if it is made at this point.

4.3.1.2 Response

Especially in light of my argument from the last two chapters, I do not want to deny that there are likely to be a great many differences between the ways two given people think. The question that is important here, though, is not simply whether there are differences at all, but is instead whether there are differences that should be characterized as differences in which concepts the two people possess—say, as opposed to differences in which beliefs the two people have. Given the theoretical role I am assigning to concepts, another way to put the question is as whether the differences between the ways the two people think are global—as opposed to there being some things that are importantly the same about those ways, in spite of their differences.

Indeed, the objection itself is phrased (perhaps somewhat uncharitably) in a way that brings out the difference between the question of whether there are differences at all and the question that is really important here. If two people think different things about a given topic, then there is something importantly the same about what they think: they are thinking about the same topic. Two people thinking different things about a given topic is importantly different from one person thinking one thing about one topic and another person thinking another thing about another topic.

If the objection is that on any realistic picture of human thought, people rarely even think *about* the same thing as one another—that on any realistic picture, there is rarely any point of contact at all between the thoughts of different people—then the objection simply does too much violence to our ordinary picture of ourselves as social, thinking creatures to be appropriate here, although it might be entirely appropriate in a debate in psychology or cognitive science.

On the other hand, if the objection is that on any realistic picture, there are many differences between the ways different people think, but these differences are not global—but there are still points of contact between the thoughts of different people—then the objection prompts the question of how to characterize whatever is the same among people. And given the role I have assigned to them—for example, determining the topic about which someone is

thinking but not typically determining what someone is thinking about that topic—concepts seem to be an excellent candidate to be the things that are truly shared among people.

4.3.1.3 Holism

Of course, someone suspicious of the distinction between features of a concept and beliefs involving that concept—someone, that is, who holds that every belief involving a given concept is a feature of that concept—does not assign quite the same theoretical role to concepts that I do, so such a person might not be satisfied with my response to this objection. But the theory that most naturally follows from denying that there is a distinction between beliefs involving a given concept and the features of that concept is a holistic inferential role theory of concepts, and as I have already pointed out, this kind of theory has some trouble when it comes to publicity.

On the one hand, if publicity is understood as many people having many of the same concepts as many other people, then as I argued explicitly in the first chapter of this work[11], individualistic holistic inferential-role theories are incompatible with anything even remotely approaching publicity. This is because they count every belief in which a concept is involved as an individuating feature of that concept, and for any concept that an individual possesses, there is very likely at least one belief in which that concept is involved that is particular to that individual.

On the other hand, if publicity is instead understood as many people having many concepts that are similar to concepts many other people have, an individualistic holistic inferential-role theory of concepts does little better. This is because, as I argued in the previous section, individualistic inferential-role theories have little hope of giving a useful account of similarity of concepts when they cannot account for many people having many of the same concepts as many other people. And the problems I outlined in that section are even worse for holistic theories, since they bring in a great many more dissimilarities among concepts, thanks to their holism.

However, it is worth again pointing out that non-individualistic (i.e. social) holistic inferential-role theories do not necessarily have any of the problems I have highlighted for individualistic holistic inferential-role theories.[12] But as they are not a target of my argument, their proponents should have no reason to object here.

4.3.2 Similarity allows for borderline cases

The objection in the last subsection was, in a way, too general. It offered no reason to think that concepts in particular will rarely be the same from one person to the next. So this

[11] And as others have argued, particularly Fodor (1998; 2004).

[12] For example, none of the complaints I have made in this work are relevant to the account of publicity that Brandom (1994) offers as part of his holistic inferential-role theory. But in my taxonomy, Brandom's is a social theory, so it is not one of my targets.

objection and those in the next two subsections are meant to be more specific and more clearly relevant to concepts.

4.3.2.1 Objection

Although he is not focused on concepts in the same way I am, Stephen Stich offers some observations in *From Folk Psychology to Cognitive Science* that are clearly relevant here, since while developing what he "take[s] to be the correct account of our folk psychological notion of belief," he claims "that *sameness* of content plays no role in our folk conception of belief. Rather, what we invoke in making judgments about how a belief is to be characterized is a tacit notion of *similarity*." (1983, 7) In justifying this claim, Stich offers two lines of reasoning that can serve as the foundations of two important objections to my argument, and so I will discuss Stich's reasoning in this subsection and the next.

The first line of reasoning Stich offers begins with a few detailed cases, real and imagined, in which the way one person thinks slowly changes—due to progressive dementia, say, or to learning more and more about some topic. As the change occurs in these cases, there is a point at which it is clearly appropriate to say that the person believes the same thing as some other person, and there is another point at which it is clearly appropriate to say that the person does not believe the same thing as some other person does. But there is no point at which the dividing line between believing the same thing and not believing the same thing clearly should be drawn—there are 'borderline' cases where it is not clearly appropriate to make either claim.

Stich then briefly describes a similar case involving his young son and concludes that *believing the same thing* is really a matter of similarity:

> How much physics must my son know before it is appropriate to say he believes that $E = mc^2$? The more the better, of course; but there are no natural lines to draw. ... What these cases suggest is that the relation I have been calling 'content-identity' is actually a *similarity* relation, one which admits of a gradation of degrees. (1983, 85-6)

Now, Stich puts his conclusion, as well as all of his cases, in terms of the content of beliefs—the p in "believes that p"—rather than in terms of concepts. But as I described in the first chapter of this work, fixing content is one of the important theoretical roles I am assigning to concepts, so anyone who reasons as Stich does would likely offer the following objection to my argument:

Objection: In the cases Stich describes, there are factors that vary by degrees, with our intuitions about concept sharing varying accordingly: on one side of a spectrum we confidently affirm concept sharing, while on the other side of that spectrum we confidently deny it. Additionally, in these cases there is no single point on the spectrum at which our intuitions flip; we do not directly move from affirming to denying concept sharing. Instead, the cases exhibit vagueness, or at least some similar kind of

indeterminacy, when it comes to our intuitions about concept sharing. As this kind of indeterminacy on a spectrum typically comes along with properties and relations that come in degrees, this suggests that in our ordinary picture, concept sharing is a relation that comes in degrees. And since similarity comes in degrees but sameness does not, this suggests that in our ordinary picture, concept sharing is a matter of similarity.

4.3.2.2 Response

Now, I certainly cannot deny Stich's descriptions of the cases he presents. It just seems obvious to me, as apparently it does to him, that there are many cases in which factors that vary by degrees can make the difference between two people thinking the same thing (which requires sharing concepts) and not thinking the same thing (because they do not share concepts). And it also seems obvious that among these are cases in which it is not clearly appropriate to affirm or to deny concept sharing—that there are borderline cases, as one typically finds with vagueness.

But it simply does not follow from these that concept sharing comes in degrees. To see this, consider an analogous argument that *being a couch* comes in degrees:

Couchness comes in degrees: We can imagine a spectrum of pieces of furniture with clear couches on the one side and clear armchairs on the other side, where each piece of furniture differs from others on the spectrum only in its width, which determines its point on the spectrum. Width is clearly a matter of degree, and this imagined spectrum shows that it can make the difference between being a couch and not being a couch. Additionally, somewhere in the middle of this spectrum, there will be some pieces of furniture that are neither clearly couches nor clearly not couches: there will be some borderline cases. This shows that being a couch is vague. Therefore, being a couch comes in degrees.

Now, it might not be obvious whether being a couch really does come in degrees or not, but I do not want to devote much space to that question. To make it at least somewhat plausible, if it is not already, that being a couch does not come in degrees, compare being a couch to width. Even if two objects are both clearly unqualifiedly wide—likely thanks in part to context determining some kind of width threshold—then because width comes in degrees, it still makes sense to compare their widths. If the two objects have determinate widths[13], then either one will be wider than the other, both will have the same width, or the other will be wider than the one. On the other hand, if two objects are clearly couches, then it makes little sense to try to compare the degrees to which they are couches. This might be because being a couch does not come in degrees.

In any case, I hope it is at least clear that this argument is not enough to show that being a couch comes in degrees. What this argument does show is that vagueness, at least

[13]Objects might have indeterminate widths because they have no clear orientation (to distinguish width from length or depth, say) or because their boundaries are somewhat indeterminate. But these complications are not relevant here.

in some cases, importantly involves some property or relation that comes in degrees but that need not be the vague property or relation itself. In this case, width comes in degrees, and being a couch in part depends on width. But because being a couch does not depend on width in a fully determinate way—because there is no precise width required for being a couch, even keeping other factors fixed—being a couch exhibits borderline cases when width is varied. This is all that is going on in this case, and it simply does not follow from couchness depending on width in an indeterminate way that couchness comes in degrees.

The question, then, is whether concept sharing is more like width—which obviously comes in degrees and which can exhibit borderline cases in part because it does—or more like being a couch—which does not obviously come in degrees but which exhibits borderline cases in part because it depends in an indeterminate way on width, which does. I think that especially given the considerations I offered above in this chapter, the most reasonable answer is that concept sharing does not itself come in degrees but that it exhibits borderline cases because concept sharing depends in an indeterminate way on factors that do come in degrees.

After all, as Stich's cases illustrate, concept sharing clearly does depend on various factors that come in degrees, including possession of related knowledge, inferential ability, and others. And it would be somewhat strange to suggest that concept sharing depends on all these factors in a fully determinate way. Thus we have an explanation for the phenomena Stich identifies that is consistent with the considerations I offered above in this chapter because it does not require that concept sharing come in degrees and so be a matter of similarity, rather than sameness. The only interesting similarity in these cases is similarity of background knowledge or inferential ability, rather than similarity of content or concepts.

4.3.3 Similarity is context-dependent

4.3.3.1 Objection

The other line of reasoning that Stich offers and that I will discuss here has to do with context dependence:

> The suggestion that content-identity (so-called) is actually a similarity relation is reinforced by the context dependence of many intuitions about how beliefs are appropriately described. ... Consider, to choose just one case, the child who accepts and repeats a few isolated sentences of a complex scientific theory. If we tell a story in which it is important that the child *assert* $E = mc^2$ or *deny* $E = mc^4$ it can seem quite natural to say that the child does indeed believe (or know) that $E = mc^2$. ... If, by contrast, we tell a story in which it is important that [the child] be able to use the belief to solve some contextually important problem, our inclination is to deny that she believes $E = mc^2$. This shift in judgment, as a function of the focus of our interest in the context where the judgment is called for, is quite characteristic of similarity judgments. If we are discussing climate and terrain and I ask, "Is the USSR more similar to Canada

> or to Cuba?" Canada is the clear intuitive choice. But if our discussion has been about political systems and I raise the same question, the natural answer is Cuba. (1983, 86)

Stich offers other cases as well to support the claim that belief ascriptions are context dependent in this way, but I hope that the claim is already fairly plausible.

Stich's reasoning could motivate the following objection to my argument:

Objection: Intuitive judgments about concept sharing—such as judgments about whether a child shares with her teacher the concepts involved in the belief that $E = mc^2$—can vary with context, as Stich's cases show. This context dependence is characteristic of judgments about relations of similarity, and it is not characteristic of judgments about relations of sameness. So judgments about concept sharing are judgments about a relation of similarity. Since these intuitive judgments are founded in our ordinary picture of ourselves, in that picture concept sharing is a matter of similarity.

4.3.3.2 Response

Now, again I do not wish to deny the data that Stich presents in his cases. And I also do not wish to deny the premise of the objection that concept sharing can vary with context.[14] But again, I think both Stich and my hypothetical objector are too quick to move from the data they offer to the conclusion that the relation in question is a matter of similarity.

The problem with this objection is that it takes context dependence to be more special than it actually is. Context dependence is common to intuitive judgments of many kinds[15] including of sameness, not only judgments of similarity. Again, consider an analogous argument about couches:

Couch identity is a kind of similarity: Suppose that my couch is damaged beyond repair and that a friend and I seek to replace it with a couch of the very same kind. We look through a furniture catalog to determine the model of the couch, my friend sees a photo of a couch with the right shape but with different upholstery and says, "This couch is the same as yours," and I agree. Then, having determined the model, we go to a store with many differently-upholstered couches of that model (including, say, one with the same upholstery as the couch in the catalog photo), but because none of those

[14] I think with this objection the move from content-identity (with which Stich is concerned) to concept sharing (with which I am concerned) is more suspicious than it was with the previous objection. For example, one could reasonably claim that there is no context in which it is appropriate to say that the child capable only of parroting the sentence "$E = mc^2$" possesses the concepts that are involved in the belief a knowledgeable physicist expresses with that sentence—one could claim that in, say, knowing that E is equal to mc^2 rather than mc^4, the child merely knows something about which of two sentences is true without knowing anything about how energy relates to mass. But I do think that there are some cases in which judgments about concept sharing can reasonably vary with context, even if this is not one of them, so I will accept this move.

[15] Travis (1985), for one, discusses how massively context-dependent many judgments of different kinds are.

> couches have the same upholstery as mine, my friend says, "None of these couches are the same as yours," and again I agree. Our behavior is, I think, entirely reasonable. This shows that intuitive judgments about whether two couches are of the same type depend on context. Therefore, couch-type identity (so-called) is actually a similarity relation.

Just as with the objection above, the conclusion simply does not follow, and indeed it seems false. There is clearly more than mere similarity at work here. If it helps, allow me to stipulate that the correctly-shaped couch in the photo really did have exactly the same shape as my couch, and that my friend only said it was the same because she (somehow) knew this.

What the couch case helps to bring out is that judgments of sameness can vary with context because two objects can be the same (truly the same) in some respects while still being different in others, and we are not always interested in the same kinds of sameness in different contexts. For example, two couches can have the same shape without having the same color, and we can be interested only in shape in some contexts but still be interested in more in other contexts. Or, more interestingly, two people can be inclined to make the *same* assertion when asked whether $E = mc^2$ or $E = mc^4$ without being inclined to make the *same* inferences from that assertion, and we can be interested only in assertions in some contexts but still be interested in more in other contexts.

The upshot of Stich's observations, then, in this subsection and the last, is that intuitive appropriateness of belief ascriptions exhibits significant indeterminacy and context dependence. And because in our ordinary picture of ourselves belief ascriptions are closely linked to judgments about concept possession and sharing, any theory of concepts intended to respect our ordinary picture should leave room for significant indeterminacy and context dependence. This lesson will become relevant in the next chapter, where I will sketch a social theory of concepts. But it importantly does not follow that respect for our ordinary picture requires understanding concept sharing as a matter of similarity.

4.3.4 Similarity is not transitive

4.3.4.1 Objection

Although again he is not talking about concepts explicitly, in *Thought* Harman offers a line of reasoning that might motivate an objection to my argument here. Harman begins his presentation of this reasoning with a brief discussion of translation:

> One general scheme of translation is better than another to the extent that it is simpler, preserves dispositions to accept sentences under analysis in response to observation, and preserves similarity in usage. Each of these desiderata is a matter of degree, and they compete with each other. ... sometimes we do give up the simplest way to translate the sentences of a friend, i.e., the identity mapping, in order to preserve more of his dispositions to accept sentences under

> analysis or to gain more similarity in actual usage. The fit is never perfect; but let us assume that there is sometimes a unique best fit: a best general scheme of translation.
>
> The fit is only a matter of degree, so the translation relation need not be transitive. If x is the translation of y and y is the translation of z, x may not be the translation of z. ... This tells against the suggestion that translation should pair sentences under analysis that have the same semantic representations ... (1973, 107-8)

The claim that a relation that is a matter of degree will not be transitive should already be familiar. And the claim that the relation of being the best translation is not a transitive relation should be uncontroversial. To see this in a toy example, imagine that L is a language incapable of describing cats. The best translation of "cat" from English into L is "a," the best translation into English of which is "small, furry animal." Translating "cat" from English to L and back, then, results in "small, furry animal," but translating "cat" directly from English to English results in "cat." Thus translation is not transitive.

One important new idea that Harman introduces here, though, is the idea of translating using 'the identity mapping.' He seems to have in mind cases where the best translation of a sentence from one person's idiolect into another person's idiolect is the exact same sentence (or perhaps a sentence that sounds or looks exactly the same). For example, if a friend says, "That is a cat," and I wish to agree with her using my own idiolect, then I will likely translate her utterance simply as "That is a cat."

Harman then makes use of this idea in giving an account of 'sameness' of meaning on which it is really a matter of similarity:

> Two people can be said to mean exactly the same thing by their words if the identity-translation works perfectly to preserve dispositions to accept sentences under analysis and actual usage. To the extent that the identity-translation does not work perfectly, people do not mean *exactly* the same thing by their words; but if the identity-translation is better than alternatives we will say that they mean the same thing by their words. Here we mean by *the same*, *roughly the same* rather than *exactly the same* ... The only sort of sameness of meaning we know is similarity in meaning, not exact sameness of meaning. (1973, 109)

My argument so far concerns concepts, not public-language words, but much of the argument has dealt with the meanings—or at least the denotations or other semantic values—of concepts, so someone who follows Harman's line of reasoning might offer the following objection:

Objection: As Harman observes, "is the best translation of" denotes a relation that is not transitive and so must be a matter of similarity, rather than sameness. And whether we are working with public-language words or with concepts, when comparing the meanings of two terms the most we can hope for is that each be the best translation

> of the other. So instead of comparing concepts by asking whether they have the same semantic value, as much of your argument had us doing, we should instead ask how they translate. If we do this, we will inevitably end up with a picture on which there is no true sameness of concepts but only similarity, and this will bring us closer to having the correct picture.

4.3.4.2 Response

I am, as I already suggested, very sympathetic to the claim that "is the best translation of" denotes a non-transitive similarity relation, rather than a sameness relation. But this is no surprise, since "is the best translation of" can clearly be analyzed in terms of being the closest fit of some kind, which is clearly a matter of similarity.

I also do not wish to deny that meaning should be analyzed in terms of translation, though my motivation in this case is a desire to remain agnostic, rather than a wish to affirm the claim in question. Harman has reasons for analyzing "meaning the same" in terms of translation, rather than in terms of semantic value, that simply go beyond the scope of this project.

But I will take issue with the more specific claim that meaning should be analyzed in terms of being the *best* translation—that the only kind of translation we might be interested in is the best translation.[16] Something like "is a perfect translation of" denotes a relation that is clearly importantly different from that denoted by "is the best translation of." In particular, while being the best translation is fairly clearly a matter of similarity, being a perfect translation is most plausibly a matter of sameness.

Another important difference between being a perfect translation and merely being the best translation is that not all terms have perfect translations into a given language, while all terms will typically have some best translation—though the best might not be very good.[17] It is this difference, I think, that lies behind the mistake in the objection above: In comparing the meanings of two terms using translation, we will often simply look for the best translation of each term into the language of the other because we are almost guaranteed to find one[18], whereas if we were looking for a perfect translation we might be looking in vain. But it does not follow that we never look for a perfect translation—that we never care about sameness of meaning in an exact sense.

---

[16] In Harman's presentation, this claim is not made explicitly. But it shows up in the way Harman moves from talk of 'a best general scheme of translation' to talk of '*the* translation' (my emphasis), a move that one might reasonably question.

[17] In other words, "the best translation in L of x" plausibly expresses a function from words to words that is total, whereas "the perfect translation in L of x" clearly expresses a function that is often partial.

[18] In some important cases, we will not be able to find the best translation. But assuming we are not simply limited by our knowledge of the languages or our ability to reason, this will be because there is no *single* best translation. For our everyday purposes in comparing meanings, though, this will not be a serious problem. We are not really interested in finding *the* best translation so much as we are interested in finding *a* best translation—a translation such that there are no better translations.

Now, Harman claims that the goodness of a translation scheme is a matter of fit and that fit will never be perfect, so he might deny that we can ever reasonably hope for a perfect translation even of a single term. But this denial would more be a consequence of Harman's particular view of translation than it would be a consequence of his metasemantic view. To take an oversimplified example: A theorist who denies that translation is ever perfect might claim that translating "and" (used to express conjunction) from one English speaker's idiolect to another English speaker's idiolect using the identity mapping will not be a perfect translation because there are many sentences involving "and" that one speaker will accept but the other will not—for instance, if there is any simple sentence that one speaker accepts but the other does not, then the one speaker will accept some conjunctions with that sentence that the other will not. There is clearly room for views of translation that give this verdict, and Harman might have one.

But we do not have to accept this verdict. There are many reasonable views of translation on which this kind of difference will not be relevant to how good the translation is. On these less strict views, we might take only some parts of the ways both speakers use "and" to be relevant to how good the translation is—that is, we might reject holism. Or we might even judge the translation by some other means than comparing the two speakers as individuals in isolation from one another—that is, we might reject individualism. Again, we can see that a kind of holistic individualism (concerning public-language words this time, rather than concepts) is antithetical to true publicity, but again, we have the option of rejecting the view rather instead of giving up on publicity.

So my hypothetical objector is wrong in claiming that the most we can hope for is that each term be the best translation of the other, although in some cases that is the most we can expect. Instead, the most we can hope for—at least if we have the right kind of theory—is that each term be a perfect translation of the other. And in my argument above, I highlighted a few kinds of cases in which we do really seem to care about perfect translation or some other matter of sameness.

Of course, my argument from the last two chapters might seem to imply that we will only very rarely be able to find a perfect translation from the language of one person into the language of another, because there are so many differences between the ways people think and use language. But that only follows if we have an individualistic theory of the terms in question. As I have suggested throughout this work, on a social theory we will be able to find many points of sameness between many different people, which will greatly enhance our prospects of finding perfect translations. And as the last two subsections suggest, on an analysis that respects our ordinary picture of ourselves as social, thinking creatures, perfect translation ought to exhibit significant indeterminacy and context dependence, so it might not be quite as demanding as it first seems.

4.4 Conclusion

So a proponent of an individualistic theory of concepts who wishes to respect our ordinary picture of ourselves will not be able to use similarity of concepts to account for something like the publicity of concepts, at least without incurring a significant cost. This is because in using similarity in this way, an individualist will be giving up on some important features of our ordinary picture. These include both transitivity of concept sharing and a distinction between possessing a technical variant of some common concept and possessing the common concept but using it incorrectly. Additionally, on an individualistic inferential-role theory of concepts, it is not even clear what similarity might be.

This concludes my argument that individualistic theories of concepts come with a significant cost because they are incompatible with publicity. The only thing I have left to do now to motivate social theories of concepts is to present a social theory, at least in outline, so that I can discuss the costs that come with it. This is what I will do in the next, final chapter.

Chapter 5

A Social Theory of Concepts

In this chapter, I will present a sketch of a social theory of concepts. It will be a rough sketch, for two reasons. First, my project is necessarily limited in scope, and in the project as a whole I am devoting more space to my negative argument that individualistic theories of concepts have a certain problem with publicity than to my positive proposal for a specific social theory. This is because in current discussions of theories of concepts, this particular problem with individualistic theories does not seem to be well-known, but there do exist well-known social theories of concepts.

The second reason I will only present a rough sketch of a social theory is that although the difference between social and individualistic theories is an important one when it comes to theories of concepts, it seems to me to be largely orthogonal to many other differences. For example, whether a theory is social or individualistic, it is possible for that theory to identify a concept in part according to certain of its causal relations, to all or part of its role in inference, or to some other feature. And just as in my discussion of individualistic theories of concepts I tried to remain agnostic about other theoretical features where practical, I will try to do the same for social theories. In this way, I hope that proponents of individualistic theories of various kinds will see the social component of my proposed theory as a potential amendment to their own theories. To be sure, this would be a significant amendment, and the resulting social theory would differ from the original individualistic theory in several important ways. As I argued in chapters 2 and 3 of this work, individualistic theories are deeply incompatible with publicity. But incompatibility can run deep without involving every component of a theory.

So my main goal here will be to present my theory. Before I do that, though, I will briefly summarize the state of play as I see it at this point in my project. My summary will focus on social theories of concepts as a family, since the last three chapters were largely focused on individualistic theories. After this summary, I will present my theory in two passes: first I will paint in broad strokes the picture of what any social theory should look like, and then I will fill in some details particular to my own theory in a more speculative way. Once the details of the theory are at least somewhat clear, I will go on to discuss some interesting consequences of and objections to the theory. And along the way, I will compare it to a

couple of closely-related theories.

5.1 The state of play

As I have said before, my main aim with this project is to motivate a social theory of concepts. A social theory seems to me to be the only kind of theory compatible with the publicity of concepts, and publicity is important if we want to use concepts to distinguish cases in which different people are genuinely communicating, agreeing, or disagreeing from cases in which two people are merely talking past one another.

5.1.1 Publicity is important

Cases in which it is not clear whether two people are talking past one another occur regularly both in everyday life and in the kind of theoretical discussions that are common in philosophy. For example, as I have already mentioned, it is often unclear whether philosophical discussions about concepts are really discussions about concepts or are instead simply two people going back and forth making claims involving different technical terms that look or sound like "concept"—say, because one person is using "concept" in a more normative way and the other is using "concept" in a more descriptive way, or because one person is using "concept" to denote a kind of abstract object and the other is using "concept" to denote a kind of mental particular.

In these cases, there is often an important distinction between, say, genuine and merely apparent disagreement. To continue the above example: If one person really is claiming that concepts are importantly normative entities and another person really is denying this, then the two disagree. Given that they disagree, it would be reasonable for them to proceed in their discussion by focusing on this disagreement and seeing whether they can resolve it instead of focusing on some other disagreement that follows from this one, such as a disagreement about what internal structure concepts must have. On the other hand, if one person is making the claim "concepts are importantly normative entities" using "concepts" as a technical term (for, say, the entities that necessarily have the features that concepts are purported to have on this person's favored theory of concepts) but recognizes that there is a straightforward way to translate "concepts" as the other person is using it into her own vocabulary (by, say, translating "concept" in the other person's vocabulary as "realizer of a concept" into her own), then the two might not genuinely disagree. If they do not disagree, then it would be reasonable for them to proceed in their discussion by settling on some more neutral terminology and then focusing on other areas of potential disagreement, rather than by focusing on this mere terminological difference. Mere terminological differences are often not very philosophically interesting.

A very natural way to account for this distinction between genuine disagreement and mere terminological difference that grounds the distinction in broader philosophical theory is as a distinction between two people having different beliefs involving the same concept and

two people having different beliefs involving different concepts. For example, if one person has the belief CONCEPTS ARE NORMATIVE and the other has the belief CONCEPTS ARE NOT NORMATIVE, then they genuinely disagree. On the other hand, if one person has the belief CONCEPTS_1 ARE NORMATIVE and the other has the belief CONCEPTS_2 ARE NORMATIVE, where CONCEPTS_1 and CONCEPTS_2 are different concepts, then they may well not disagree.[1]

In this way, an interest in the distinction between genuine and merely apparent communication, agreement, and disagreement naturally leads to an interest in individuating concepts.[2] But a theory that individuates concepts will not be helpful when it comes to this distinction if on that theory it is very rare for two different people to share a concept, even a high-level concept like CONCEPT or CAT. That is, a theory will only be helpful when it comes to this distinction if on that theory concepts are public, including high-level concepts like CONCEPT or CAT. This is because if concept sharing is practically required for communication, agreement, and disagreement, then if concepts are not public nearly every case of potential communication, agreement, and disagreement (even cases that seem obviously to feature these phenomena) will turn out to be a case in which different people are merely talking past one another because they turn out not to share the relevant concepts.

So publicity of concepts is important. But as I argued over the last three chapters, individualistic theories of concepts are incompatible with publicity. This is because if an individualistic theory were to identify the possession conditions for a complex concept—possession conditions that distinguish the concept from other similar and similarly complex concepts—then because of the many subtle differences among thinkers, it will turn out to be extremely unlikely that any given thinker will possess the concept. So whatever complex concept someone expresses with a word like "concept" or "cat," on an individualistic theory it will be extremely unlikely that her interlocutors will share that concept.

This problem with individualistic theories, then, leads us to social theories.

5.1.2 Social theories can achieve publicity

In contrast to individualistic theories of concepts, which simply identify possession conditions for a given concept that an individual thinker either will or will not meet, social theories treat concepts as importantly social entities. In particular, social theories take concepts to have social features—such as by whom they are shared—that are more basic than other

[1] Whether a difference in concepts is sufficient for disagreement, of course, depends on what disagreement is. As I pointed out in the first chapter, there are different things theorists can mean by "disagreement" just as there are different things they can mean by "concept." But on most interesting analyses of disagreement I have encountered—including those offered by MacFarlane (2007) in a very helpful discussion of disagreement—a difference in concepts rules out disagreement in a typical case, even if not in every possible case.

[2] If concepts are abstract entities, as I am treating them, then the interesting question is how to individuate them. On the other hand, if they are mental particulars, then the interesting question is how to type them. At least for my purposes in this project, this difference is a perfect example of a merely terminological difference.

individuating features such as semantic values or inferential roles. So in a case like the one described above, in which the question arises of whether two individuals are expressing the same concept when they use the word "concept," a social theory would not answer the question by first determining which concept the one thinker possesses, second and separately determining which concept the other thinker possesses, and then finally giving a verdict as to whether they are the same concept or two different concepts, as an individualistic theory would. Instead, a social theory would have us first take up the question of whether the two thinkers share a concept or not before taking up the question of which concept it is (or which concepts they are).

One advantage of this approach—the advantage that is relevant here—is that it does not result in concept sharing depending on two thinkers being similar in the many ways that would be required for them to each separately possess a given complex concept. Instead, the two thinkers might share a concept in virtue of some other, easier-to-come-by commonality or other relation. To give a significantly oversimplified example, on a social theory it might be that the two thinkers share a concept in virtue of shared membership in a certain community—say, the community of speakers of English, or the community of participants in the contemporary academic philosophical discussion of concepts—and intentions to speak and think as members of that community.

On individualistic theories of concepts, the difficulty—indeed, the impossibility—with which I have been most concerned is that of determining possession conditions for a given complex concept that are both demanding enough to distinguish the concept from the many possible concepts that are similar but not the same and also lax enough that many different thinkers can hope to meet them. These possession conditions must be so demanding because on an individualistic theory they do the work of individuating the concept, distinguishing it from all other concepts an individual could possibly possess.[3] But on any theory for which publicity is a desideratum, these conditions must also be lax because a diverse group of thinkers must be able to meet them. However, since a social theory does not assign the same kind of importance to possession conditions, at least as they are applied to individual thinkers, it does not have to wrestle with this tension.

On a social theory, then, an individual can possess a concept simply by being a member of a certain community or by being related to another individual in a certain way. Because of this, on a social theory there is hope that many individuals will share many concepts, even complex concepts like CONCEPT and CAT. But because these kinds of possession conditions are not solely responsible for identifying or individuating a concept, their relative laxity is not problematic—or at least it is not problematic in the way that it would be on an individualistic theory.

Instead, on a social theory the difficulty comes in explaining how this kind of community membership or interpersonal relation is sufficient for concept possession, particularly when that membership or relation must itself be explanatorily prior to the publicity of concepts.

[3]Peacocke (1995) makes this point very explicitly, but it is implicit in the work of other individualistic theorists.

I will discuss this difficulty in the next section, since it is an important feature common to all social theories.

5.2 Social theories of concepts

The feature that characterizes social theories of concepts as a family is that determining the community of thinkers who share a concept is prior to determining other features of the concept, including semantic features. As a consequence, if a concept is identified in some way that leaves open exactly which concept it is—for example, if a concept is identified as whichever concept a certain speaker expresses with the word "concept"—and the question arises of exactly which concept it is, a social theory would have us answer with a story very different from one an individualistic theory might have us tell.

In this section, I will present an outline of this story. I will not go into much detail, instead focusing on the big picture, particularly on the differences between social theories and individualistic theories as families of theories of concepts. In this way, this section will serve as a point of contrast with the second chapter of this work, in which I focused on the way an individualistic theory might tell this story without going into the details that distinguish one particular individualistic theory from another.

5.2.1 Social theorists face an explanatory dilemma

On a social theory, the story of exactly which concept a given concept is begins with the determination of who shares the concept. Since this comes before determination of the semantic value of the concept, the sharing of the concept must not depend on the concept having any particular semantic value. In the previous chapters, I made this point in several discussions by presenting cases in which the semantic value of one thinker's concept depended on their sharing the concept with another thinker, such as an expert. Since that style of case should be familiar by now, I will now make this point in a somewhat different way: I will present a dilemma for the social theorist, highlighting the difficulty that is particular to a social theory of concepts.

So then: Suppose we are interested in which concept a certain speaker is expressing with the word "concept"—in particular, we are interested in whether this speaker is expressing the general concept CONCEPT or she is expressing some more idiosyncratic concept. If we are trying to develop a social theory of concepts, our first step will be to identify with whom this concept is shared, perhaps by identifying the community of language users of which the speaker is speaking and thinking as a member. If in using "concept" she is speaking and thinking simply as a member of the broader English-speaking community or the broader philosophical community, say, then she may well be expressing CONCEPT, whereas if she is speaking and thinking as a participant in some debate in which "concept" is used as a somewhat technical term, then she may well be expressing a more idiosyncratic concept with it.

What, though, grounds the existence of these communities and determines membership in them? In accounting for these communities—these seemingly intentional social entities—the social theorist faces a dilemma: whether to attempt to reduce these social entities to individualistic entities or to take social entities as explanatorily primitive.

5.2.1.1 Social reduction and regress

On the one hand, a social theorist might attempt to account for these communities and other social entities in terms of individualistic entities.

For example, membership in the communities in question very plausibly depends, at least in part, on the intentions of the individuals who might make up that community. So to continue our case from above: if someone intends to use "concept" to express a non-technical concept but speaks in a room full of others using "concept" in a technical way, the intention to speak and think as a member of the broader linguistic community can plausibly override the context that might otherwise grant the speaker membership in the narrower community. So in identifying the community of which our speaker is a member, we might reasonably look to her intentions.

If we did so, while taking this kind of community membership to determine concept sharing, we might include the following claim in our theory:

intention determines sharing: For any concept C, a thinker shares C with all other members of a community x if there is some term t such that in using t to express C the thinker intends to express the concept that other members of x (*qua* members of x) express with t.

Now, there are at least several clearly serious problems with this claim that might be considered problems of oversimplification. For one thing, it does not leave open the possibility of such an intention failing, which certainly seems to be a live possibility if, say, the thinker in question is deeply mistaken about how t is used or what its semantic value is. For another, it is far from obvious that we often have intentions quite like these or how these communities could become objects of these intentions.

But the problem with this claim that I want to press is a kind of regress problem. If this is how our theory works, then in order to determine whether a speaker is expressing CONCEPT, we will first have to determine whether the speaker intends to express the same concept as do members of some CONCEPT-possessing community. Assuming the speaker can think of that CONCEPT-possessing community as THAT COMMUNITY, this requires the thinker to have the intention EXPRESS THE SAME CONCEPT AS DO MEMBERS OF THAT COMMUNITY—and having this intention requires the speaker to possess the concept CONCEPT.

So using this claim, our explanation of a speaker's sharing of the concept CONCEPT depends on her already possessing CONCEPT. This is not yet a regress problem in the strictest sense, because we might explain the speaker's possessing CONCEPT in some other way. However, we will very clearly have a regress problem if we strengthen the claim:

intention determines possession: For any concept C, a thinker possessing C consists in there being some community x and some term t such that members of x (*qua* members of x) express C with t and the thinker, in using t, intending to express the concept that members of x (*qua* members of x) express with t.

If having the relevant intention requires possessing CONCEPT, as it very plausibly does, then with this claim the explanation of any speaker possessing any concept quickly reaches an inescapable regress.

For example, according to this claim, in order to possess CAT a thinker must (among other things) intend to express the concept that members of some (in fact CAT-possessing) community express with some term, but intending this requires possessing CONCEPT; and in order to possess CONCEPT a thinker must (among other things) intend to express the concept that members of some (in fact CONCEPT-possessing) community express with some term, but intending this requires possessing CONCEPT; and so on.

The difficulty, then, for a social theorist who wishes to reduce the social to the individualistic comes in the tension between accounting for concept possession as a kind of community membership, which is at the heart of social theories as a family; and accounting for this community membership as a matter that is in some way intentional and so conceptual, which is clearly very intuitive. If concept possession depends on membership in a community—or if concept possession depends on bearing some other relation to some other social entity, such as playing a social game or being subject to a social norm—which in turn depends on concept possession, then a regress will be difficult to avoid.

And it is somewhat hard to see how a theorist who wishes to reduce the social to the individualistic could avoid claiming that the social entities that ground concept possession—communities, games, or norms, say—depend in some way on concept possession. After all, if these truly are *social* entities, then surely they must arise out of our existence as social creatures—rather than, say, arising more directly out of our existence as conglomerations of physical particles. And if these entities are sophisticated enough to settle complex matters like concept possession, then surely they must arise out of our existence as thinking social creatures, since other aspects of our social existence will not feature the necessary sophistication.[4]

So for any social theorist who wishes to account for the social in terms of the individualistic, explanatory regress seems inescapable.

5.2.1.2 Social primitivity

On the other hand, a social theorist might make use of explanatorily primitive social entities in accounting for conceptual communities or other social entities involved in accounting for shared concepts.

[4]These are matters I will return to in the next subsection, where I will discuss the importance of public language.

These primitive social entities might be the communities themselves. Or they might be the social norms that sustain the communities—determining membership, among other things—with the communities being accounted for in terms of these primitive social norms. Or they might be still other social entities, with the communities being accounted for in terms of certain social norms that are themselves ultimately accounted for in terms of these other, primitive social entities. Yet another option would be to claim that the social norms that sustain these communities depend on other social norms which in turn depend on other social norms, with the explanation being social and normative 'all the way down'—a potentially infinite chain of social entities that, taken as a whole, is in effect primitive.

To be sure, there are some good reasons to adopt this strategy. Chief among them is the difficulty of taking the other horn of the dilemma—of somehow reducing the social to the individualistic. In this work, the last three chapters showed the impossibility of accounting for a certain social phenomenon—the publicity of concepts—from within a certain individualistic framework. In the above subsubsection, I identified a serious problem for any theorist who attempts this social reduction in particular. And we are all likely familiar with examples from other works of attempted social reductions that are ultimately unsatisfactory.

In addition to the difficulty of doing otherwise, a social theorist might adopt this strategy in order to give an account that is social through and through, an account on which concepts and related entities are not just social but necessarily social—an account on which concepts simply are not the kinds of things that could arise from an arrangement of purely individualistic entities, no matter how complicated that arrangement is. This seems to be a large part of what motivates Brandom (1994), in giving his social theory of concepts, to make a claim of primitive social normativity in giving his account of the rules of what he calls "the game of giving and asking for reasons," rules which he takes to ground concept sharing.

However, a claim of social primitivity will not sit well with all theorists. For example, although they fall ouside the scope of this work, I have metaphysical motivations for reducing the social to the individualistic generally. So I take it to be a desideratum for any theory that it does not introduce any explanatorily or metaphysically primitive social entities.[5]

Additionally, a theorist attempting to account for shared concepts in terms of some other social entities and to then claim that those other social entities are primitive is likely to encounter a problem similar to the regress I identified in the above subsubsection. For example, if a theorist claims that concept sharing depends on a certain kind of community membership that is governed by certain primitive social norms, then if those norms take that community membership to depend on the intentions of the thinkers in question and so on the concepts possessed by those thinkers, there will still be an explanatory regress.

Of course, a social theorist who encounters this regress and who has already made use of primitive social entities might simply avoid the regress by claiming that concepts themselves

[5]I recognize that the project of reducing the social to the individualistic is not universally shared, though, and I certainly do not take anything I am saying here to be a serious objection to social theories of concepts that take this horn of the explanatory dilemma.

are primitive social entities. But this would prevent that theorist from giving satisfactory answers to many of the questions that motivate this project, such as how we can use concepts to distinguish two people truly disagreeing from two people merely talking past one another.

I, then, prefer the first horn of the dilemma, and I will discuss my solution to the explanatory regress later in this chapter. But my approach to this dilemma is not integral to the rest of my theory, so I will separate that discussion from the core presentation of my theory in the following section, presenting it later in section 5.5. And readers who are sympathetic to social theories of concepts and who prefer this second horn will have no difficulty seeing how to adapt my theory to use social primitivity instead of a social reduction: much like Brandom, I take concept sharing to be grounded in the game we play when we treat ourselves and others as social, thinking, linguistic creatures; and so much like Brandom, if I preferred to give a theory that was social through and through, I would claim that the rules of this game either are or are grounded in primitive social norms.

5.2.2 Public language is important to concept sharing

The other feature I will discuss that I think should be common to all reasonable social theories of concepts is the role assigned to public language. By "public language" I mean to refer to languages like English that are public in that they are available to be used by many people. As with concepts, though, I will not take this publicity for granted.

5.2.2.1 Why public language

As should be familiar by now, the distinguishing claim of a social theory is that concept sharing consists in, is sustained by, is realized by, or is otherwise explained by some interpersonal relation that is explanatorily prior to concept identity. In giving a social theory, one of a theorist's main tasks is to identify this relation. But it is important to keep in mind that this must not be a relation that holds or does not hold between two thinkers[6] unqualifiedly. Instead, this relation must be able to hold or not hold between two thinkers with respect to a certain concept the one thinker possesses and a certain concept the other thinker possesses, because it is possible for two thinkers to share some of the concepts they possess without sharing others.

So the kind of relation a social theorist should be looking for is really an interpersonal concept-concept relation. This rules out coarser-grained person-level relations like *intending to cooperate* or *being of the same species*. And as I showed over the last three chapters, abstract relations among concepts like *being causally connected in the right way to the same property* or *playing similar roles in inference* also will not do: mere qualitative sameness or similarity of one feature or another of the concepts themselves either will not be enough or

[6]In this subsection, I will focus on the concept sharing relation as it holds or does not hold between two people. It is worth keeping in mind, though, that the publicity of concepts must involve concepts being shared among more than just two people, so the sharing relation must somehow be able to hold among many people, either in an indirect way thanks to transitivity or in some more direct way.

will be too much for sharing. What other kinds of relations does this leave, then, as good candidates for the social theorist's sharing relation?

If the relation is to be concept-specific but is not simply to amount to sameness or similarity of conceptual features, it is hard to see what it could be other than some kind of concrete relation—such as a causal relation—between the concepts as they are possessed by the thinkers in question.[7] For if the relation is abstract then it must be a relation that holds or does not hold in virtue of some features of the concepts themselves, rather than in virtue of the ways the concepts are possessed by the thinkers in question. So any kind of abstract sameness or similarity is out, for all the reasons I covered in the last three chapters. And although there are a great many abstract relations that are not some kind of sameness or similarity, it is very hard to see how any could be a better candidate than sameness for grounding concept sharing—how could two concepts be shared in virtue of being different in some particular way?

Of course, this is far from an airtight argument that concept sharing absolutely must amount to a concrete relation, but giving such an argument would require a great deal of space, and my space is limited. In this section I am more interested in giving a broad description of a typical reasonable social theory than in giving a broad description that must cover absolutely any possible social theory. Indeed, this entire chapter is meant to be much more speculative than the preceding chapters. So in the interest of covering more ground, so to speak, I will keep my discussion here fairly loose.

So the sharing relation will be concrete and will be concept-specific, rather than holding or not holding unqualifiedly between two thinkers. Since there are many interpersonal concrete relations, this might seem to leave many candidates. But again, some candidates are better than others, especially because many interpersonal concrete relations will not be concept-specific in the way required to ground concept sharing. For example, if two people are both intentionally interacting with a single object in their shared environment—and so interacting with each other—there is no guarantee that they will share a concept.

Simple mutual interaction with a single object cannot ground concept sharing because whichever concepts the two thinkers are using, they themselves need not enter into the interaction itself. Thus, even though the interaction must involve concepts because the two thinkers are each interacting with the object intentionally, there is an important sense in which the interaction itself is not conceptual. On the other hand, linguistic interaction—which as I am understanding it is something richer than two people each making sounds or other signs that merely causally affect the other's behavior—is importantly conceptual. This is because important features of the interaction itself, such as whether it is successful or not, can depend on how the concepts it involves relate to one another.

[7]If concepts are abstract entities, as I am treating them in this work, this will likely be a concrete relation between the concept or concepts' realizers or tokens in the minds of the two thinkers. On the other hand, if concepts are mental particulars, and so sharing concepts is a matter of having concepts of the same type, then this will be a concrete relation between the concepts themselves. Perhaps a more general way to describe the relation is as a relation that can hold or not hold in part in virtue of the features of the two thinkers in virtue of which they possess the concept or concepts in question.

There may well be other concrete interpersonal relations, beyond linguistic relations, that are also conceptual relations in this way. For example, at least on some understandings, joint attention and joint action involve a kind of mental contact among different people that is plausibly conceptual. But however complicated these other relations might be, it is very hard to see how they could approach the breadth and sophistication of linguistic relations. And if the social theorist is looking for a relation to ground the widespread sharing of many kinds of concepts that our ordinary picture of ourselves involves, she will need a relation with this kind of breadth and sophistication. After all, how could something like joint attention or joint action ground the sharing of concepts, including concepts as abstract as PRIME NUMBER and as nebulous as BEAUTY, between thinkers as different as Aristotle and a modern-day grade-school student?[8]

And if the relation is linguistic in the relevant sense, then it must involve linguistic entities that are somehow public, simply because these entities must somehow relate different people to one another. Whether they are possible or not, private languages will be of little use here.

So if the social theorist's sharing relation is to be concrete and also concept-specific in the right way, there seems to be no better candidate than some kind of linguistic relation—or some kind of linguistic relation supplemented by other relations. In developing my social theory, then, I will focus on a concept sharing relation in which public language plays an important role.

5.2.2.2 Other theorists' use of public language

I hope I am not doing anything very contentious in attaching this importance to public language. While there are surely other broad strategies for accounting for the publicity of concepts, using public language seems to be a particularly popular one.

Indeed, most of the attempts to account for the publicity of concepts that I discussed in chapter 3 of this work involve an important role for public language. In attempting to explain how the content of a concept possessed by an ignorant person can match the content of a concept possessed by an expert, Harman (1982) suggests that the ignorant person's concept derives its content from a word as the ignorant person uses it, which in turn derives its content from a same word as the expert uses it, which in turn derives its content from the expert's concept: a concept-word-word-concept relation. Fodor (1995) attempts to solve the same problem in a similar way, by suggesting that an ignorant person's concept will carry the same information as an expert's concept if the ignorant person consults with or otherwise defers to the expert in using the concept. Fodor is not explicit about the role of public language in this account, but it should be fairly clear that much actual expert consultation—and much of this kind of deference more generally—will be unavoidably linguistic.

The other attempt at solving something like the problem of publicity that I focused on in chapter 3 is that of Peacocke (1995). Peacocke, like Harman, explicitly brings public

[8]This is not to say that there is no role for relations like joint attention or joint action in grounding concept sharing. But if they have a role, it will be a supplementary one—linguistic relations will still have to do much of the heavy lifting.

language in to the heart of his attempt. However, Peacocke's attempted solution and his specific problem are somewhat different from those of Harman and Fodor. Perhaps the differences of greatest relevance to this discussion are that Peacocke relates the ignorant person to a linguistic community, rather than to a single expert language user, and that he relates the ignorant person to the community via a single word that both use, rather than (as in Harman's attempt) via some kind of connection between a word as used by one and a word as used by the other.

However, one of the shortcomings of Peacocke's attempt to account for thinkers using concepts without mastering them is that he takes the publicity of public language for granted. His proposal requires that there be a single word that an ignorant thinker and members of a wider linguistic community all use, but he does not account for the semantic value (or some other kind of meaning) of this public-language word, and he does not explain how all these people are able to use this word when they do not all possess the concept it expresses. This is why his proposal cannot distinguish between experts—thinkers who are using a word that shares a semantic value with a concept they possess—and ignorant thinkers—thinkers who are using a word that does not share a semantic value with any concept they possess. I discussed this problem in more detail in chapter 3. One lesson to draw from Peacocke's attempt, then, is that while public language is potentially very useful in accounting for the publicity of concepts, the publicity of public language must not be taken for granted.

Of course, I discussed Fodor, Peacocke, and Harman in chapter 3 as examples of individualistic theorists, so they might not be as relevant to this discussion as social theorists accounting for the publicity of concepts. Unfortunately, I am aware of very few social theories of concepts that even approach the thoroughness and level of detail of Fodor's and Peacocke's theories. Those social theories I am familiar with, though, do attach similar if not greater importance to public language.

I will not discuss any social theories in detail in this section, because I wish to compare my theory to other actual and possible social theories as I present it. Instead, I will just briefly mention two social theories and the roles they assign to public language. Perhaps the only social theory of concepts that is as well-worked-out as Fodor's and Peacocke's theories is the theory presented by Brandom (1994), and in Brandom's theory there is little if anything that distinguishes concepts and public-language words. Another interesting theory is that offered by Fine (2007), and although it is not clear that Fine's theory is a social theory or a theory of concepts, strictly speaking, Fine does present a relation very like the social theorist's concept sharing relation, and he thus relates thoughts to one another interpersonally via language.

5.3 The theory

One upshot of the last section is that at the heart of a typical reasonable social theory there will be an interpersonal concept sharing relation that is explanatorily prior to concept identity and that importantly involves public language. And for a given concept possessed by a given individual, the theoretical role of this relation is to determine the extent of the

community of thinkers with whom that individual shares that concept.[9] But within these constraints, there is significant room for the details of one social theory to vary from the details of another.

This section is about the details of the social theory of concepts I am offering as part of this project. In the following subsections, I will discuss the aims of the theory—what *kind* of theory it is, in the sense of the last chapter—and then I will move on to the details of the concept sharing relation itself, including the role of public language. I also wish to discuss my solution to the explanatory dilemma I presented in the last section, but I will leave this discussion to a later section, since in some ways it is less central to my theory than, say, the role of public language.

Before I get to the details of the theory, though, I want to repeat my earlier disclaimer: this discussion will only offer a sketch of the theory, and a fairly rough sketch at that. As I have said before in this work, my main aim is to motivate social theories of concepts generally when compared with individualistic theories. This is different from the aim of motivating one particular social theory when compared with other social theories. With respect to my main aim, this chapter is meant to be a proof of concept, showing that social theories of concepts are achievable and giving some idea of the costs associated with social theories. Once the costs associated with social theories are apparent, they can be compared with the costs associated with individualistic theories—particularly the costs I discussed over the last three chapters related to the publicity of concepts. In this section, then, as in the last, the discussion will be fairly loose.

5.3.1 This is a folk psychological theory

Philosophical thought experiments can do many things. They can highlight the importance of certain theoretical implications, and they can help to locate the exact point at which two theorists' views diverge. But they are perhaps most commonly used to access our everyday intuitions, to check the answer our 'folk' theories give to some question or other.

For example, 'Twin Earth'-style thought experiments[10] might show—among many other things—that concepts like GOLD and WATER are importantly different from concepts like SHINY YELLOW METAL and CLEAR NOURISHING LIQUID in that the former concepts somehow aim at natural kinds in a way that the latter concepts do not. But if the experiments show this, they do not do so by explicitly arguing for some theory and working through the implications of that theory until this claim about natural kind concepts is reached. Instead, in the experiments, cases are described to the reader, and the reader is asked to make certain judgments about the cases. These judgments are taken to be implications of some theory or other that the reader has already adopted, typically an implicit folk theory.

In light of those results, a philosophical theory of concepts that is intended to systematize our folk theory or otherwise respect our ordinary picture of ourselves will do better if it

[9]This is not to say that every concept every individual possesses is shared with others. As I will discuss later, the sharing relation need not be instantiated for some concepts.

[10]for example, those offered by Putnam (1973)

accounts for this way that certain concepts aim at natural kinds than if it does not account for this. And if such a philosophical theory is incompatible with concepts aiming at natural kinds in this way, it comes with a significant cost. So even though the specific nature of natural kind concepts might not be as central to a theory of concepts as, say, the metaphysical status or the internal structure of concepts, a proponent of a given theory of concepts has some reason to show how the theory can account for the way that some concepts aim at natural kinds. Although there were likely others, this is very plausibly one consideration motivating Laurence and Margolis (2002) to show how a theory of concepts like Fodor's might account for this natural-kind aim.

This kind of story should, of course, be very familiar, both from philosophy generally and from this work specifically. Throughout this work, I have offered thought experiments, from Burge's arthritis case (Burge, 1979) to some of my own, that are meant to draw out the intuition that many concepts somehow aim at being shared—just as Twin Earth cases and others draw out the intuition that many concepts somehow aim at denoting a natural kind. I have shown how certain theories are incompatible with the widespread sharing of concepts and so come with a cost. And in this chapter, I will show how the theory I favor can account for this sharing.

The theory I am offering, then, is not intended as a revisionist theory. I will not be accounting for the publicity of concepts simply because doing so offers theoretical benefits like simplicity or productivity—though I think it clearly does offer these benefits, and its doing so tells strongly in its favor. Instead, in developing this theory I am trying to take the implicit folk theories that drive our intuitions when we read Burge's arthritis case and make those theories more explicit and systematic—and, where necessary, to fill in the gaps and smooth over the inconsistencies in those theories. In the terms I used in the last chapter, I will be giving a folk psychological theory. To put it in the terms I will use going forward: rather than coming up with a new game we might start playing, I want to codify the rules that are implicit in the game we are already playing when we treat ourselves and others as social, thinking creatures.[11]

5.3.2 Many concepts are public-language terms

As I suggested, public language plays a central role in my account of the the publicity of concepts. Indeed, the role is so central that on my theory many concepts are the same terms as public-language words—CAT is one and the same term as "cat," for example. In the following two subsections I will explain why this is and what it means.

5.3.2.1 The same word relation

Before I explain what it is for a certain concept to be the same term as a certain public-language word, I should say a bit about what it is for different uses of a public-language

[11] Brandom (1994) makes similar use of game-rule talk, and I am following him in adopting that talk.

word to be uses of the same word. That is, I should try to make clear in what sense I am using "same word," or how I am conceiving of the 'same word' relation.

At least within a given mode—focusing only on spoken language, say, or only on written language—there is some sense of "same word" in which qualitative sameness of certain features, such as being composed of the same phonemes or the same letters in the same order, is both necessary and sufficient for two uses to be uses of the same word. But being the same word in this sense is clearly not sufficient for any kind of semantic relation, and so is clearly not sufficient for grounding, say, agreement or disagreement.[12] I am interested in grounded phenomena like these, so it should come as no surprise that I am not using "same word" in this sense.

Instead, I am conceiving of the same word relation in something like the way that David Kaplan seems to be conceiving of the relation in "Words":

> On my conception, there is a single word "color" spelled one way in Canada: "c", "o", "l", "o", "u", "r" and another way in the United States, "c", "o", "l", "o", "r". Similarly, there is a single word which is pronounced *shedge-yule* in Canada, and *skedge-oo-ul* in the United States. ... We are, of course, familiar with dialectal variation in pronunciation, but I call to your attention that there is also dialectal variation in spelling. This dialectal variation in spelling of the word "color" for example does not, repeat *not*, make for different words(1990, 98-9)

There is an important connection between "color" as written in the United States and "colour" as written in Canada, between "schedule" as spoken in the United States and "schedule" as spoken in Canada. This is the connection I am interested in. And importantly, it does not only connect the written uses of a word to one another, nor does it only connect the spoken uses of a word to one another. That is, the same word relation is not only an intramodal relation.

Thus, a single English word can be both spoken and written—not to mention that it can also be signed, recorded as a pattern of bumps on a page, transmitted as a digital pattern of electric current or light in a cable, and used in other language modes as well. And even within a given mode, as Kaplan points out, uses of a single word can take significantly different forms, such as when English speakers with different accents pronounce the word or when there are different accepted spellings of the word. And although it should be obvious, it is perhaps worth noting that there is no necessary connection among the various ways a word can be used: the world could easily have been such that the sound "cat" and the written sign "cat" were ways of using different words.

So when it comes to the question of whether two uses are uses of the same word—whether two word tokens are of the same word type—the answer cannot simply depend on whether some abstract relation, such as a similarity relation, holds between the two uses.

[12]It also cannot reliably be applied across modes with an eye to necessity or sufficiency for any kind of semantic relation.

This is not to deny that some degree of similarity of some features is required for two uses in the same mode to be uses of the same word, at least in typical cases. And it is not to deny that there are any important structural relations between, say, the ways English words are represented as marks on a page and the ways they are represented when stored and transmitted digitally (since in these modes, English words are typically composed of characters in the same ways). But even these abstract relations are clearly not sufficient for two uses being uses of the same word, and each by itself is not even necessary—for example, in a language without an alphabet, one would have less reason to expect the same structural relations between the ways words are written and the ways they are stored and transmitted digitally as one tends to find in English.

So just as with a concept used by one person and a concept used by another person, when it comes to a word used in one mode (or by one person) and a word used in another mode (or by another person), the question of whether they are one and the same depends on more than just similarity, qualitative sameness, or some other abstract relation. The 'same word' relation, like the social theorist's concept sharing relation, must be concrete in some way.

And there are other important similarities between the same word relation and the concept sharing relation as well. One worth noting is that much as the concept sharing relation must be explanatorily prior to other conceptual features, including semantic value, the same word relation must be importantly presemantic. To see this, imagine a case in which by coincidence the only people capable of using the spoken form of "cat" falsely believe that naturally having fur is part of being a cat, but that there are many experts who use "cat" in other modes who are well aware of the existence of naturally furless cats and whose use of "cat" reflects this, and the ignorant "cat" speakers act as normal ignorant speakers do—they would be somewhat inclined to defer to experts were the experts around and were the possibility of non-furry cats relevant, but they have not yet learned of their mistake. Now, in this case, consider the question of whether the spoken form of "cat" denotes *cathood* or only denotes *naturally-furry-cat-hood*.

Intuitively, of course, "cat" when spoken denotes *cathood* just as "cat" does when used in other modes. But if we were to attend only to the spoken uses of "cat," we might give the other answer, because those uses are more consistent with "cat" denoting *naturally-furry-cat-hood*. So the semantic value of "cat" when spoken depends at least in part on the uses of "cat" in other modes, and the obvious explanation of this is that someone speaking "cat" is using the same word as someone writing or signing "cat." In deciding the semantic value of "cat" when spoken, then, we must first decide whether it is the same word as "cat" used in various other modes, because the same word relation is in some way prior to semantics. This kind of story should be familiar from various discussions of concepts in previous chapters.

One last similarity between the concept sharing relation and the same word relation that I will mention here is that the same word relation must importantly involve public language—though this is surely a trivial point.

So there are several important similarities between the same word relation and the concept sharing relation. However, the same word relation is not obviously explanatorily prior or

better-understood than the concept sharing relation, so I do not wish to claim that the concept sharing relation consists in the (more basic) same word relation or that a social theorist can account for the concept sharing relation by explaining that it is the same word relation.

Instead, I claim that the two relations are really one and the same—or at least that the concept sharing relation is one and the same as the being-able-to-use-the-same-word relation, which is so closely related to the same word relation that I will not distinguish between them going forward—at the same explanatory level. One thinker's public concept uses are related to her public-language utterances and to another person's public concept uses just as spoken words in one public language are related to written words in the same public language: each are uses of the same terms in different modes.

An example, then: Imagine that person A has the thought CATS MAKE GOOD PETS and expresses the thought by writing "Cats make good pets;" and person B reads this, has the thought CATS DO NOT MAKE GOOD PETS, and expresses the thought by saying aloud "Cats do not make good pets." I claim that in this case there is a single term denoting *cathood* that was used four times in four different modes: once in A's thought, once in written English, once in B's thought, and once in spoken English. And if person C were to come along, have the thought MY CAT MAKES A GOOD PET, and (say) express that thought in American Sign Language, I claim that this term would then have been used six times in six different modes. This is in part because the lexicon of ASL is (at least in this case, according to my understanding) so closely related to that of English that this single term exists in both languages.[13] It is also because a single term can be used in grammatically different ways, such as when a single term is used with a different numbers as "cats" in one instance and "cat" in another—or such as when a single verb is used with different conjugations as "make" in one instance and "makes" in another.

So the social theorist should account for the concept sharing relation by examining what the same word relation consists in—or rather by examining the relation that holds between different uses of a concept or publicly-used word across different modes, since both the concept- and publicly-used-word-specific relations are really just instances of this more general relation. And this more general relation should hold not only between CAT as possessed by one thinker and CAT as possessed by another thinker, nor only between "cat" as spoken and "cat" as written, but among all these four uses of that term.

5.3.2.2 The same term rule

As I explained in the previous section, the concept sharing relation lies at the heart of any social theory of concepts. Now that I have identified the concept sharing relation as the same term relation, I am in a good position to discuss some of the conditions in virtue of which the relation holds.

[13] I will discuss term identity across languages more below.

And as I explained in the previous subsection, in giving my theory I intend simply to codify the rules of the game we are already playing. So in claiming that the concept sharing relation is the same word relation, I am in effect claiming that the game we play when we treat each other as social, thinking creatures is one and the same as the game we play when we interact with each other linguistically, or at least that those games share a central rule.[14] So to account for the concept sharing relation, I will offer an analysis of this rule, which I will call the 'same term' rule to distinguish it from any rules that are somehow specific to concepts or to public-language words.

The same term rule, then, is the rule that grounds the same term relation, determining when and how the relation is instantiated, so determining which pairs of uses are pairs of uses of the same term and which are not. It is thus functionally similar to the rules of football that ground the 'same team' relation, determining which pairs of players are on the same team and which are not. However, where the rules of football (at least within certain leagues) are often codified quite explicitly, the rules of the game we play in which we treat ourselves and others as social, thinking, linguistic creatures are almost entirely implicit in the behavior and thoughts of us players.

There are many methods one could use to study a rule that is implicit in a game some subjects are playing: one could ask the subjects to make explicit their own understandings of the rule, one could observe the subjects playing the game in a natural setting and construct models consistent with the subjects' behavior, or one could perform experiments while having the subjects play the game in a laboratory setting to test various hypotheses of the rule. A truly rigorous study might involve all of these methods. However, this is not the place for such a rigorous study, so in examining the same term rule I will resort to the standard philosophical method of constructing imaginary cases and reasoning about theoretical implications in order to consult my own intuitions about the rule. This method is much more convenient than the others mentioned, but in many cases its results are less generalizable. As I will discuss later, though, I do not claim that my implicit understanding of the rule is exactly the same as all others', so the lack of generalizability is less of a problem in this case than it might be in others.

So: what are the conditions in which we treat different people as using or not using the same term, be it a concept or a public-language word? It seems to me that speaking very broadly, some degree of some kind of similarity in usage, something like an intention to use the term in some kind of public way, and some kind of causal connection are relevant—that at least in a typical case, each of these three features is necessary and all together they are sufficient for using the same term as someone else. My analysis of the same term rule, then,

[14]The identification of the conceptual and linguistic games is at the heart of the theory Brandom (1994) offers. However, there are important differences between Brandom's theory and the one I am developing here. I will compare my theory with Brandom's at other points later in this chapter, but it is perhaps worth mentioning now that I do not claim, as Brandom does, that concepts are necessarily social, linguistic entities. Instead, I am claiming that the concept sharing relation, on which the publicity of concepts depends, is necessarily social and linguistic. The difference between these claims will become more clear in a later section of this chapter.

will focus on these three features. I will discuss each of them in turn.

First, similarity in usage—or at least a lack of a certain kind of confident divergence in usage: In a case in which two English speakers use the sound "cat," for example, whether they are using the same term in using that sound can vary with the degree of a certain kind of similarity in their usage. On the one hand, if their patterns of use are not quite the same but fall within the normal range of ways people use "cat" to talk about cats which I have outlined in other parts of this work and with which we are likely all familiar, then as long as no unexpected details are added to the case, we will treat the two speakers as both using the same term. On the other hand, if one speaker's use pattern falls within that cat-related range but the other speaker consistently makes claims like "cats don't run if you don't put fuel in them" and "I had to replace a tire on my cat the other day" and never makes claims like "cats meow" or "cats make good pets," then we will not treat them as both using the same term, even if they meet and (perhaps briefly) treat themselves as using the same term as one another.

Next, an intention to use the term publicly in a certain way: In a case in which two English speakers use the sound "cat" somewhat differently, whether they are using the same term can vary with the way their behavior shows or does not show an intention to use a public term in using that sound. If one speaker uses the sound as I do when I talk about cats while the other speaker uses the sound in the same way except for saying things like "part of being a cat is naturally having fur," and this other speaker says these things come what may, *no matter how much apparent disagreement they encounter*, then we will not treat them as using the same term. On the other hand, if the second speaker says these unusual things about natural furriness but eventually defers to the community's usage (or would eventually defer to the community's usage but just happens to never encounter enough apparent disagreement), then we will take that (potential) deference to show that the second speaker had something like an intention to use a public term or to use whatever term she used in a public way, and so we will treat the two speakers as using the same term—and as having used the same term from the beginning, before the eventual deference.

Before I move on to the third feature that plays into the same term rule, I should note that in distinguishing between using a public concept and using an idiosyncratic concept, in some cases it can be enough to look to this relatively coarse-grained distinction between having something like an intention to use a concept publicly and not having such an intention; but other cases require a finer-grained distinction. For example, in academic debates featuring "concept," it can be unclear whether two theorists are using the same word to express the same concept or whether they are instead talking past one another and using different concepts. But if two theorists are talking past one another in this way, it may well be that neither is using an entirely idiosyncratic concept—instead, it might be that both are using widely shared concepts that simply happen respectively to be shared in different communities, such as a more philosophical community and a more psychological community. This kind of case, in which concepts are 'public' within some communities without being shared between those communities, suggests that the publicity- or sharing-related intentions that play into the same term rule can be somewhat complicated. I will not examine this

complication here, but it will become relevant in the next section.

Last, causal connection: In a case in which one speaker uses a term and another speaker uses a term, and the first speaker uses the first term similarly to how the second speaker uses the second term, whether the terms the two speakers are using are really the same term can vary with whether and how they are causally connected. For example, if some speaker lost to history (before the days of English, even) used a term very similarly to the way some ordinary present-day English speaker uses "justice," then as long as no unexpected details are added to the case, we would treat the two speakers as both using the same term, even if the sounds or other signs the two speakers use are quite different. On the other hand, if an extraterrestrial (or a human who has had no human contact but who nevertheless develops a spoken language and who soliloquizes a great deal) uses some term very similarly to the way the present-day English speaker uses "justice," then we would treat the two speakers as not using the same term but as using merely very similar terms—the relevant difference being the lack of the right kind of causal connection.[15]

So at least upon a very brief examination, when we are playing the game in which we treat ourselves and others as social, thinking, linguistic creatures, we tend to count two people as using the same term when and only when

1. their patterns of use are similar in certain ways to at least a certain degree,

2. they seem to intend to use terms publicly in a certain way, and

3. they are causally connected in a certain way.

This analysis is, of course, less than exhaustive in several ways. For one thing, the considerations I offered in the previous four paragraphs were offered too quickly to establish in any rigorous way that each of these factors is indeed in itself relevant at all—that none of these features merely serves as an indicator for some other feature that is more directly relevant. For another, I have done nothing to establish exactly *how* each of these factors is relevant—that each is truly individually necessary, say, or that all three are truly jointly sufficient. (Indeed, I expect that there are at least some other necessary factors, though many of these will be 'background' or 'enabling' factors that not clearly relevant to the present discussion.) And lastly, I have not been at all specific in describing these factors.

I am offering this analysis, then, more as a sketchy proposal than as a triumphant conclusion.[16] I will leave it up to future work to establish more details of the same term rule in

[15]This case might seem to turn on the publicity- or sharing-related intentions of the speakers involved, but it need not. To see this, add the detail that all three speakers intend to use their terms in a public way, that they all intend to discuss universal issues relevant to all beings, that they all see themselves as part of a discussion that must be going on throughout the ages and across the universe. At least according to my intuitions, even the addition of this detail does not change the fact that the causally unconnected speaker is using a different term, simply because this speaker is not connected to the discussion of justice that has a long history in western or world culture.

[16]In doing so I at least stand in good company: neither Brandom (1994) nor Fine (2007), who both propose theories in which relations similar to the same term relation play an important role, manages to

this game I am playing. However, it is somewhat appropriate that I only offer this sketchy analysis, since I do not believe that all of the details of this rule can be made precise in public language, and I can see no guarantee that the rule by which I am playing exactly matches the rules by which others are playing—though I do expect a very high degree of similarity. I will explain both of these points in the the penultimate section of this chapter, section 5.5. And before that in section 5.4, I will discuss some interesting consequences and applications of my proposed theory, at which point at least some more details of the rule will become clear.

5.4 Consequences, applications, solutions

My presentation so far has focused more on the overall structure of my theory than on the theoretical machinery responsible for giving verdicts about, say, what concepts a given individual possesses—machinery that includes the same term rule as well as whatever machinery assigns conceptual features like semantic values. This is in part because I am more interested in giving a proof of concept of a social theory of concepts, showing that it is possible to give a coherent social theory of concepts that works within the constraints I am concerned about, than I am in working out all the precise details of such a theory.

But another reason that I have not discussed at least some important details of my theory is that I am more interested in the differences between social and individualistic theories of concepts than I am in the similarities, and as I said in the introduction to this chapter, I think social theories can share many important components with individualistic theories. For example, a social theory could share much of the machinery responsible for assigning semantic values to concepts with Fodor's theory or Peacocke's theory—the main difference being that where Fodor's theory identifies a concept's semantic value as the property to which an individual's uses of the concept are related by the right kind of causal law, a social theory could identify the semantic value as the property to which uses of the concept by any thinker who possesses it are related by the right kind of causal law; and where Peacocke's theory identifies a concept's semantic value with an eye to preserving truth in the inferences in the thought of an individual that individuate the concept, a social theory could identify a concept's semantic value with an eye to preserving truth in the individuating inferences in the thought of any individual who possesses the concept.

The only complication particular to social theories comes in potentially needing to distinguish between mistaken uses of a concept by a thinker who has not mastered the concept and appropriate uses of the concept by a thinker who has mastered it or between non-individuating inferences involving a concept made by a thinker who has not mastered the concept and individuating inferences involving the concept made by a thinker who has mas-

offer an exhaustive, rigorous analysis of his theory's relation, including of the possible conditions in which the relation holds and does not hold. They do, of course, discuss their respective relations thoroughly, but for their purposes as well as for mine, the precise details of the relation are not as important as various features of the role the relation plays in the larger theory.

tered it—but the problem of distinguishing between mistaken and appropriate uses of a concept, or between non-individuating and individuating inferences involving a concept, is a problem familiar to individualistic theorists, and adding a social dimension to the problem does not overly complicate it.

I am not interested in taking a side in the debate between causal theories, inferential-role theories, and other families of theories of concepts as to what the semantic value or other meaning of a concept consists in. This debate cuts across the debate between social and individualistic theories of concepts. So I will leave open how exactly the semantic value of a concept should be determined on my theory.

However, in spite of all the details of my theory I am leaving unspecified, there are some consequences and applications of the theory, as well as some potential problems for the theory, that are worth discussing here. In discussing these, I will try to give a sense of the social structure of concepts that I expect to emerge according to the theory, both more broadly across society and more specifically in some potentially tricky cases, and along the way I will also try to give a bit more detail to my analysis of the same term rule and relation that lie at the heart of the theory.

5.4.1 Within a linguistic community, many concepts are shared as public-language words

Most of the concepts that philosophers discuss—concepts like WATER, KNOWLEDGE, RED, and GOOD—intuitively can be and are expressed in public language—using words like "water," "knowledge," "red," and "good" in English-language philosophy, for example. This suggests that as possessed by any given anglophone philosopher, these concepts are related by the same term relation to concepts possessed by a great many other anglophone philosophers, at least according to the rules of the game we play when we treat ourselves and others as social, thinking, linguistic creatures.

And a cursory inspection reveals that the standard requirements for the same term relation are, indeed, met in most of these cases. For example, in using the concept she expresses with "water" while, say, considering a Twin-Earth-style case, a philosopher will very likely be using a concept that she uses similarly to ways other philosophers use certain concepts, that she intends to use in publicly in some way or at least to share with other philosophers, and that is causally connected in the right way (which can be very loose and indirect) to concepts as used by other philosophers. And if these requirements are met for the philosopher as related to anglophone non-philosophers, then the concept as she uses it will stand in the same term relation to concepts as used by anglophone non-philosophers as well.

So it turns out that many of the philosophically interesting terms in the conceptual idiolect of a given philosopher are each, on my theory, simply one form of a term that can be found in different forms in the conceptual idiolects of other philosophers and, indeed, of other thinkers as well. This is one way to account for the widespread sharing of these concepts on my theory.

But these concepts do not only stand in the same term relation to other concepts. They also often bear the relation to the public-language words used to express them. For example, there are important similarities in the ways WATER and "water" are typically used, WATER is often used with the intention to denote the same thing "water" denotes and to otherwise use the same term—and vice versa—and the two are, of course, hardly causally isolated from one another. The typical relation between WATER as possessed by one anglophone thinker and "water" as used by that thinker shares with the relation between WATER as possessed by one thinker and WATER as possessed by another thinker those features that enable the two thinkers to share the concept.

So another way to account for the widespread sharing of a concept on my theory is to explain how the concept as possessed by one thinker is the same term (in a different form) as some public-language word used by that thinker, which is the same term (in the same form) as that public-language word used by many other thinkers, which is the same term (in a different form) as concepts possessed by those thinkers. Thanks to the transitivity of identity, the concept as possessed by the first thinker is thus the same term as the concepts possessed by the other thinkers.

And because conceptual interaction among thinkers is so often linguistically mediated, this latter concept-word-word-concept style of account is in some ways more explanatorily useful than the more direct concept-concept account. Once it is clear how concepts are shared as public-language words on my theory, one can begin to see how a thinker as unconcerned with and unaware of the mental states of others as a young child might acquire a shared concept, how a newly-minted or otherwise private concept could become shared, and how a complex and arbitrary concept like CAT or even JUSTICE could be shared across a community as large as the English-speaking world.

This is, of course, not to say that all concepts are public-language words or that concepts are only shared as public-language words. But public language is crucial to the publicity of concepts, especially to the publicity of the concepts philosophers seem to be most interested in.

5.4.2 Some concepts are shared across languages

Within a given linguistic community, such as the anglophone community, my theory offers a relatively simple story explaining how many concepts are shared. But it might be less clear how concepts could be shared across languages on the theory. For example, given the differences between "cat" (in English) and "chat" (in French), one might wonder whether a monolingual English speaker could count as sharing the concept CAT with a monolingual French speaker on the theory.

Before I discuss "cat" and "chat," I first want to discuss another term and make the point that in our ordinary picture of ourselves, not all concepts are shared across linguistic communities. The widely-used Danish word "hygge" has a meaning close to that of "cozy" but, I am told by reliable sources, has no exact translation in English—hygge has more to do with warmth and with socializing than coziness does, in addition to other subtle

differences. Since Danes presumably think about hygge in addition to speaking about it, the concept HYGGE is presumably widely shared in Denmark. But since the term has no exact translation in English and has not been borrowed into the language by many English speakers, on my theory there is no obvious way to account for HYGGE being widely shared across the Danish-speaking and English-speaking communities. This, though, is entirely appropriate, since a lack of sharing is the intuitive verdict here: we do not judge anglophones capable of thinking about hygge things as such, at least before they have been introduced to the term when learning about Danish culture, even though we do take them to be capable of thinking about coziness, which is very similar.

Supposing that Danish has no exact translation of "cozy" just as English has no exact translation of "hygge," COZY and HYGGE are a pair of very similar concepts that are nonetheless importantly distinct, with many thinkers possessing one of the concepts without possessing the other. This phenomenon should be somewhat familiar from my earlier discussion of individualistic theories of concepts, because the picture that emerges on those theories is one on which individuals typically possess concepts that are very similar to but nonetheless distinct from concepts that other individuals possess. The difference here is that we are comparing large linguistic communities, rather than individuals, but the result is the same: without the connection provided by shared language, thanks in part to the arbitrariness of concepts, we see subtle differences in concepts rather than sharing.

Given the arbitrariness of "cat" and "chat," we should not be too surprised or disappointed if, on a theory, it turns out that they are subtly different terms, used to express subtly different concepts, rather than simply being different forms of the same term. That is, it should not be too surprising if after close study we came to the conclusion that the English-speaking and French-speaking communities respectively possess the subtly different concepts CAT and CHAT, much as the English-speaking and Danish-speaking communities respectively possess the subtly different concepts COZY and HYGGE.

But this would still be somewhat unexpected. Intuitively, "cat" and "chat" are perfect translations of one another and are used to express the same concept. What, then, is the difference between the pair "cat" and "chat" and the pair "cozy" and "hygge" that leads to this difference in the intuitive verdicts about them? Perhaps whatever leads to the difference in intuitive verdicts will also lead to a difference in the verdicts the theory gives.

There are, of course, many differences: English and French are related somewhat differently than English and Danish are; "cat" and "chat" are obviously natural kind terms, whereas "cozy" and "hygge" are not obviously natural-kind terms. But perhaps the most obvious difference is that "cat" and "chat" resemble one another much more on the page than do "cozy" and "hygge." Why is this?

I am not a linguistic historian, but my understanding is that "cat" and "chat" resemble one another because they have a common Latin ancestor, as do many other English-French cognate terms. "Cat" and "chat" both look the way they do in part because of their relations to "catus." And these relations amount to, among other things, the same term relation, for reasons that should be familiar by now. So because they are both the same term as "catus," they are both the same term as each other—and they are typically both used to express the

same concept, CAT.

The "cat"/"chat" story offered by my theory is thus fairly straightforward. However, the theory should not rely on common ancestors like "catus" in every case. For one thing, it is not clear that knowledge of "catus" or even supposition of a common ancestor is driving the intuitive verdict about "cat" and "chat." And for another, many English terms with Germanic roots intuitively have exact translations in French even though they do not have the same similarity of form resulting from such recent common ancestry.

But even though the connections between English and French are not always as straightforward as that between "cat" and "chat," the languages are connected in many ways beyond their terms with common Latin roots. Many of their terms have common Indo-European roots, and others still have important historical connections simply because there has been so much historical interaction and overlap between the English-speaking and French-speaking communities. The languages have grown into their current forms together, and this allows many pairs of English and French terms that are not as similar-looking as "cat" and "chat" to nonetheless stand in the same term relation to one another.[17]

5.4.3 Sharing is not always straightforward

In many cases, having a common ancestor or some other historical connection helps to sustain the same term relation between two forms of a term. But in other cases, terms can share a common ancestor without obviously being the same term as one another. For example, it is my understanding that "port" in English (used literally to refer to an opening, as on the side of a ship) and "porte" in French (used literally to refer to a door, as on the front of a house) share a Latin ancestor in "porta" and yet are importantly different terms with importantly different meanings today.

The existence of pairs of terms like "port" and "porte" is not clearly incompatible with the existence of pairs of terms like "cat" and "chat," though. If the way a sign is used changes enough over time, it can come to be used for a different term than it once was used for. "Port" and "porte" have changed enough in the way they are used that neither is the same term as "porta" even though each is importantly connected to "porta."

The "port"/"porte" case is, on the whole, unproblematic for the theory I am proposing—apart, perhaps, from questions it raises concerning transitivity that I will address in the next subsection. It is unproblematic in part because "port" and "porte" both clearly stand on their own as distinct terms; neither is truly dependent on "porta," even though both are importantly connected to it. There are, though, cases that are potentially more problematic, and in this subsection I will discuss two.

[17] I can imagine a scenario in which greater interaction between the English- and Danish-speaking communities leads to the terms "cozy" and "hygge" growing together so that they come to share a meaning—come to be two forms of the same term. This scenario illustrates how words in different languages without a common ancestor might still stand in the same term relation to one another, but it also raises the question of how to think about the apparent historical differences between the words—a question very similar to the one I take up in the next subsection.

5.4.3.1 Kripke's puzzle about belief

Especially given the close connections between French and English I made so much of in the last subsection, it is very plausible that "London" and "Londres" are different forms of the same term much as "cat" and "chat" are. As such, we would expect that according to my theory, these words are used to express the same concept, which we might call LONDON. The claim that there is no distinction in thought between what "London" expresses and what "Londres" expresses, though, is at the heart of a famous puzzle posed by Kripke—a puzzle that might seem to make trouble for my theory.

Kripke (1979) asks readers to suppose that thanks in part to learning the two languages separately and so not realizing that "Londres" and "London" refer to the same city, Pierre comes to have both a belief that he expresses in French as "Londres est jolie" and a belief that he expresses in English as "London is not pretty." Assuming that "jolie" and "pretty" express the same concept, as do "Londres" and "London," the upshot seems to be that Pierre can have the belief LONDON IS PRETTY and the belief LONDON IS NOT PRETTY and yet not realize that they are contradictory—even if he considers both beliefs together and reasons blamelessly.

Kripke's case is, of course, puzzling. It might seem to motivate a tightening up of the requirements of the same term rule so that "Londres" and "London" as Pierre uses them will not count as the same term because, say, he uses them too dissimilarly. But this would be a mistake. The ways Pierre uses these words might be no more dissimilar than the ways two English speakers use "London," and tightening up the requirements in this way would do more harm to the theory than good for it.

The treatment of the case that is most compatible with my theory—and, I think, that is most reasonable generally—involves recognizing that as Pierre uses them, "Londres," "London," and the concept each expresses all corefer because they are all different forms of the same term. This, indeed, is why the puzzle is so puzzling. But just as Pierre is unaware of the relation between "Londres" and "London," so is he unaware of the relation between the concept he expresses with "Londres" and the concept he expresses with "London." He is unaware that these concepts are really a single concept—a concept he shares with himself in much the same way that a typical French speaker who has heard of London shares a concept of the city with a typical English speaker who has heard of it.

If anything, rather than causing a problem, the positing of a close relationship between concepts and public-language words and the recognition of the possibility of possessing concepts without mastering them help my theory to deal with this puzzle. To anyone with a practical familiarity with public language, it should be uncontroversial that a speaker can use different forms of a single term without any awareness of doing so—for example, if she does not realize that two sounds are merely alternate pronunciations of a single term. Kripke's case simply shows that it is similarly possible for a thinker to use different 'forms' of a single concept without any awareness of doing so—an awareness the thinker would likely have if she had fully mastered the concept.

However, in addition to the "London"/"Londres" case, Kripke offers the "Paderewski"/

"Paderewski" case which is structurally similar but in which the 'two' words do not differ in their spelling or pronunciation. As Kaplan (1990) points out, in light of the possibility of this case it is difficult to see how features like spelling or pronunciation could be directly relevant to the questions of word individuation that the "London"/"Londres" case raises. And upon consideration, it is similarly difficult to see how analogous conceptual features such as mode of presentation could be directly relevant to the questions of concept individuation that the "London"/"Londres" case raises. So the real explanation of the "London"/"Londres" case cannot be that Pierre somehow possesses two 'forms' of a single concept, since the form itself is not directly relevant.[18]

Instead of a difference in concepts or a difference in forms taken by a single concept, what relevantly distinguishes what Pierre possesses that he expresses with "London" from what Pierre possesses that he expresses with "Londres" must somehow be a difference in the possessions. In the one case he possesses the concept LONDON in one way, and in the other case he possesses the concept LONDON in another way, where these ways of possessing are not features of the concept but are instead features of Pierre's relation to the concept.

There are, of course, many features of Pierre's relations to LONDON that might seem to distinguish what he expresses with "London" from what he expresses with "Londres"—and unfortunately, I cannot offer a detailed and well-motivated account of what distinguishes Pierre's relations to LONDON from those of a speaker who does not share Pierre's confusion. Instead, I can only gesture in the direction of an account, leaving the details and motivations for another work.

What seems to me to be the relevant difference between Pierre's uses of "London" and his uses of "Londres" is that the former uses are related by something like what Kaplan (1990) calls "intrapersonal continuity" to when Pierre acquired the concept LONDON from uses of "London," whereas the latter uses are similarly related to when Pierre acquired the concept LONDON from uses of "Londres." This intrapersonal continuity need not amount to any kind of conscious awareness on Pierre's part, but it is nonetheless real, and Pierre's two tracks of LONDON-involving intrapersonal continuity distinguish him from others who do not share his confusion about London.

Or to make what I think is the same point in less temporal and more social terms: There are two groups of features of Pierre in virtue of which he could be said to possess LONDON—to be a member of the LONDON-possessing community. On the one side, there are the features in virtue of which "London" as Pierre uses it to express LONDON bears the same term relation to "London" as used by many anglophones to refer to London. On the other side, there are the features in virtue of which "Londres" as Pierre uses it to express LONDON bears the same term relation to "Londres" as used by many francophones to refer to the same city. We can see that these are separate groups of features of Pierre by considering that he can, for example, change his "London"-regarding intentions without changing his "Londres"-regarding intentions—or by considering the difference in the causal

[18]For somewhat similar reasons that lie beyond the scope of this work, the real explanation also cannot simply be that Pierre is not aware that his different uses of LONDON are really uses of the same concept.

ancestries of his uses of "London" and "Londres" that I highlighted above.

These sketchy observations likely raise more questions than they answer, but I hope that they can give some sense of how a social theorist might hope to deal with Kripke's puzzle. At the very least, I hope that they go some way to showing that a social theorist is no worse off than an individualistic theorist when it comes to Kripke's puzzle—though that might not be saying much.

Before I move on, it is perhaps worth mentioning that even if there is an important distinction between the ways Pierre possesses LONDON, there is ample room on my theory for cases that are like Kripke's case but also like the "port"/"porte" case in that the original term produces two descendents that are distinct terms because they have enough of an existence independent from the original and each other and are dissimilar enough from the original and each other in their use.[19]

5.4.3.2 Technical and dialectal variation

The other potentially problematic case I will discuss is far more common but usually more subtle than cases like Kripke's. It is the case of a term that has a technical or dialectal variant that shares some but not all of its features—that is distinct from but in some ways dependent on the original term.

For example, it is my experience that in some contexts—in some dialects of American English, really—"barbecue" can be used literally to refer to any meat that has been cooked in a certain way, while in other contexts "barbecue" can only be used literally to refer to pork that has been cooked in that way—and it is no coincidence that the two uses allow exactly the same cooking methods.[20] The pork-specific use is parasitic on the more general use (or perhaps vice versa), although the uses are importantly distinct.

Because the two uses share the cooking-method-related part of their meaning, it would be a mistake to treat them as uses of two entirely independent terms. To do so would be to

[19]Kripke's puzzle seems to motivate Fine (2007) to claim that his counterpart of the same term relation is not transitive, but only in cases like the one in the puzzle in which a term appears to be shared—or 'coordinated,' as Fine puts it—both between and 'within' speakers or thinkers. This is because Fine wishes to allow for a case in which term uses are coordinated within a speaker as in the "London"/"Londres" case but are not coordinated between certain speakers, more like in the "port"/"porte" case. This seems to me to be a serious misstep, both because it introduces a distinction between between- and within-speaker term sharing that undermines some of the social motivations of the theory and because it makes analysis of large webs of term sharing very ungainly. Additionally, it does not seem to me that Fine is in any way forced to this conclusion, because there is ample room on his theory to allow both for cases of the "London"/"Londres" kind and for cases of the "port"/"porte" kind as long as they are carefully distinguished—though it must be acknowledged that there will be cases that are indeterminate between the two kinds.

It seems to me that Fine might be led to this conclusion in part because he takes Kripke's puzzle to show that there can be something like partial sharing or partial coordination, although again Fine could acknowledge the existence of partial coordination while maintaining that 'full' coordination is transitive. I will offer my own treatment of partial sharing in the next subsubsection.

[20]There are, of course, other contexts in which "barbecue" can only be used in other ways still, but I am only interested in these two uses.

invite complaints similar to those I made in chapters 2 and 3 of this work, because intuitively the cooking methods allowed by the two uses really are exactly the same, but thanks to the arbitrariness of cooking methods we would not expect two entirely independent terms to require exactly the same cooking methods. It would also be a mistake to treat the two uses as uses of the same term, because there really are contexts in which 'barbecue' encompasses meats of all kinds and contexts in which 'barbecue' is only pork.[21]

One potential option would be to treat the two uses as uses of complex terms: in one context "barbecue" means the same as "a meat cooked by barbecuing" and in another "barbecue" means the same as "pork cooked by barbecuing." This treatment would seem to allow for the kind of partial distinctness that the case calls for, but it too has drawbacks. For one thing, given that users of "barbecue" typically do not explicitly define the term they are using, treating the term as if it has such a strict definition might seem somewhat ad hoc. Also, relatedly, there are many cases like that of "barbecue" but in which the partial sharing of meaning is more subtle and therefore less amenable to this treatment because such clear definitions are not forthcoming. For example, when comparing the ways "concept" is used in different philosophical contexts, one is likely to find important differences and important samenesses, but although it is often possible to identify the differences—such as whether the term refers to abstract or concrete objects, or whether it covers general concepts only or leaves room for concepts that refer to individuals—it is very difficult to identify all the points sameness and explicitly describe them in simpler terms.

This kind of case, then, suggests that although the same term relation might be of great importance in accounting for the publicity of concepts and public-language words, the complexity of publicity demands an additional relation in its account, which we might call the 'variant term' relation. This relation is more complex than the same term relation because it does not simply relate or not relate terms as used by two different people or in two different forms. Instead, when the variant term relation is instantiated, it must include some specification of what differs between the two terms. So for example, I grew up using "barbecue" to refer only to pork, and when I encounter someone using "barbecue" to refer to meats of various kinds, I might treat that person as using a term that is the same as "barbecue" as I use it *except that it can refer to meat of any kind, not just pork*—and such a person encountering me might treat me as using a term that is the same as "barbecue" as she uses it *except that it can only refer to pork*.

The variant term relation, then, is somewhat more complex than the same term relation but admits of a similar analysis: it is sustained by a rule in the game we play in which we treat ourselves and others as social, thinking, linguistic creatures. Indeed, once the variant term relation has been admitted into the theory, the same term relation can be seen as

[21] And it would not be the most natural treatment of this case to take a contextualist approach: to try to account for the difference between contexts in which "barbecue" can refer to any meat and contexts in which "barbecue" can only refer to pork by treating "barbecue" as a single monolithic term that is simply used in different contexts—such as contexts in which many kinds of meats are relevant and contexts in which only pork is relevant. For one thing, "barbecue" can be used in the pork-only way in contexts in which other meats (including other meats cooked by barbecuing) are just as relevant as pork.

simply a default case of the variant term relation in which there is no difference between the used terms it relates—and it is important to the analysis that the variant term rule defaults to no difference, since that both respects our ordinary practice and is in giving an account of widespread full sharing of concepts.

5.4.4 The same term relation is transitive

When discussing the "port"/"porte" case above, I claimed it was not problematic for my theory. At the very least, though, the case raises an important question concerning the transitivity of the same term relation. And as I have pointed out at several points, it is important that the same term relation be transitive, not least because that transitivity is central to my account of the sharing of complex concepts between thinkers who are not directly acquainted with one another.

As I understand their history, the English "port" and the French "porte" are both descended from the Latin "porta." It seems likely, or it is at least possible, that there was no precise point in history when "porta" spawned the clearly distinct term "port" or the clearly distinct term "porte"—that there was, instead, a gradual change in the use of "porta" in diverging populations that ultimately resulted in the two terms we have today, which are clearly distinct from one another.

This possibility suggests the following challenge to the transitivity of the same term relation:

> Consider the term "porta" that was in use at a point in time, say, two thousand years ago. Since that point, it is possible linguistic change has been gradual enough that no one using the term "porta" or its descendents was ever using a term that is not the same term as a term in use a year ago. That is, it is possible that for any year n between now and two thousand years ago, for anyone using "porta" or one of its descendents in the year n, someone in the year $n-1$ was using the same term. So anyone using "porta" or one of its descendents in 2014 is using the same term as a term that was in use in 2013, and so on. But if the same term relation is transitive, then it follows that anyone using "porta" or one of its descendents in 2014 is using the same term as a term that was in use two thousand years ago. And since "port" and "porte" are both descendents of "porta," it follows that anyone using either of those terms is using the same term as a term that was in use two thousand years ago. But supposing that only the one term "porta" was in use two thousand years ago, it follows that "port" and "porte" as they are used today are the same term as one another. This, of course, is false, since the two terms have importantly different meanings. Therefore, the so-called 'same term' relation must not be transitive.

This challenge is worth taking seriously. But its structure should be familiar from other areas in philosophy. In particular, the structure should be familiar from discussions of

vagueness and from discussions of the identity of persons and certain other objects over time. In both of those areas, cases are discussed in which the accumulation of apparently insignificant differences reaches an apparently contradictory significance.

So if "being the same term as" is a vague two-place predicate, or if the identity of terms over time is something like the identity of persons or ships over time, then this challenge will not be so troubling. That is because when challenges like this are raised in discussions of vagueness, we should not conclude that vague predicates have no real application; and when challenges like this are raised in discussions of personal identity, we should not conclude that we can never rely on the transitivity of identity of persons or ships in our reasoning. Instead, we should simply recognize that there are some strange scenarios in which we should be cautious in applying vague predicates and in relying on the transitivity of identity of persons and ships.

And, of course, we have every reason to expect "being the same term as" to be vague and to expect the identity of terms over time to be similar to the identity of persons and ships over time. Terms are exactly the kind of hard-to-pin-down socially constructed entities that give rise to vagueness and these puzzles concerning identity. After all, the alternative would be for terms to act like numbers when it came to identity—but "being the same term as" is obviously much more complex than "being the same number as," and as I discussed in chapter 2, it is precisely this complexity that gives rise to arbitrariness and so to vagueness.

So the "port"/"porte" case does show that there are limits to how far we can rely on the transitivity of the same term relation in accounting for the sharing of complex concepts: if we try to link two terms today via some two-thousand-year-old common ancestor, we run the risk of going too far. But most cases in which we are interested in concept sharing—certainly most of the cases I have discussed in this work—are cases in which thinkers are much more closely connected, even when they are still indirectly connected. And the "port"/"porte" case does not show that we cannot rely on the transitivity of the same term relation in reasoning about those cases.

5.4.5 The same term rule enables the robust publicity of thoughts

In the first chapter of this work, I focused on the goal of accounting for the publicity of thoughts. As part of that discussion, I distinguished between two kinds of publicity of thought, which I called 'minimal' and 'robust' publicity. I also introduced the publicity of concepts as an important necessary condition for the publicity of concepts. Now that I have outlined my strategy for accounting for the publicity of concepts, it is worth returning to the publicity of thoughts and to the minimal/robust distinction to see how my account relates to those topics.

First, it should be clear that with this account of the publicity of concepts, a serious barrier to accounting for the publicity of thoughts has been removed. Even if a theorist is only concerned with the publicity of thoughts in the minimal sense involving many people actually

sharing many thoughts with many other people, for all the reasons I gave in the last three chapters of this work, as long as that theorist held that thoughts were composed of concepts then that theorist would not be able to account for the publicity of thoughts without adopting a social theory of concepts such as the one I am giving in this chapter. This is because if thoughts are composed of concepts, then sharing thoughts requires sharing concepts[22], and so the publicity of thoughts (widespread thought sharing) requires the publicity of concepts (widespread concept sharing), and so individualistic theories of concepts are incompatible with the publicity of thoughts even in this minimal sense.

This is, of course, not to say that all barriers to accounting for the publicity of thoughts have been removed. There is almost certainly more to sharing thoughts than merely sharing the concepts those thoughts involve. For one thing, sharing thoughts plausibly involves respectively having thoughts that share a certain structure that is importantly distinct from the matter of which concepts they involve. Additionally, there are questions related to indexicality and context sensitivity that come up when comparing thoughts across individuals that might not seem as relevant when comparing concepts across individuals—or at least that have not come up in my discussion of concepts here, because I have been concerned with other questions.[23] However, these other barriers to accounting for the publicity of thoughts fall outside the scope of this project.

But as I suggested in chapter 1, some theorists hoping to account for the publicity of thoughts might have in mind a more substantial phenomenon than the mere widespread sharing of thoughts. In our ordinary picture of ourselves as social, thinking creatures, there might be more to the publicity of thoughts than mere widespread sharing. This is why I expressed a desire to account not only for what I called 'minimal' publicity but also for what I called 'robust' publicity of thoughts: I hoped to offer a theory capable of doing justice to the full richness of our ordinary picture of ourselves.

As I described it in chapter 1, robust publicity of thoughts involves thoughts being shared widely, transitively, among large groups of people, easily, in spite of the many differences among the sharers, and thanks to social interaction. And now that I have given my account, I hope that upon a little reflection it will be clear that on the account, concept sharing has all of these features—especially as I have already discussed some of the features quite explicitly. That is, I hope that it will be clear that on the account, concepts are robustly public.

I also hope that upon a little reflection it will be clear that in accounting for the robust publicity of concepts, I have removed a serious barrier to accounting for the robust publicity of thoughts. But even more than that, the social theory of concepts I am offering offers at least a partial explanation of why the robust publicity of thoughts has the features it does. For example, as I explained in chapter 1, on our ordinary picture of ourselves, sharing thoughts (which is involved in phenomena like communication, agreement, and disagreement) is in some cases enabled by social interaction—a layperson previously incapable of having

[22]I discussed the relationship between thoughts and concepts in greater detail in chapter 1, so my reasoning here is somewhat quick.

[23]These questions of indexicality and context sensitivity show up in discussions offered by MacFarlane (2007) and Madagáin (2011), for example.

the thought that water is composed of hydrogen and oxygen, say, might gain the ability to have that thought by interacting with an expert who is discussing the composition of water. My theory explains why the sharing of thoughts is enabled by social interaction because my theory explains why the sharing of concepts is enabled by social interaction, and in many cases that sharing of concepts in turn enables the sharing of thoughts—such as when a layperson acquires the concepts HYDROGEN and OXYGEN from an expert, thereby gaining the ability to think that water is composed of hydrogen and oxygen.

Lastly, in chapter 1 I claimed that even though minimal publicity involves less than robust publicity, the way to account for minimal publicity is by accounting for robust publicity. This claim, too, has by now been shown to be true, since I showed in the intervening chapters that individualistic theories of concepts are incompatible with minimal publicity, and in this chapter I showed how the only alternative—a social theory of concepts—brings robustness with it. Thus, even if a theorist is not interested in robust publicity for its own sake, an interest in minimal publicity will likely lead to an interest in robust publicity.

5.4.6 Some concepts might not be expressed publicly

Another question that a social theory ought to answer—a question I already mentioned above—is the question of whether or not concepts are necessarily social in some important sense. This is closely related to the similarly important question of whether or not there could be concepts that are private in some important sense. These questions arise for my theory in particular, in part because of the way my theory accounts for the publicity of concepts. If, as my theory claims, concepts are shared by way of public language, then is there room for 'private' concepts that are not expressed in public language? If so, how do these concepts relate to concepts that *are* expressed in public language?

My answer to these questions is complicated enough that I will devote a full section of this chapter—the section immediately following this one—to answering one version of them. But before I do that, I want to point out that the answers to the questions depend heavily on the senses of "social" and "private" we take to be important.

One sense in which a concept could be private, and so fail to be social, is that that concept could in fact only be possessed by a single individual. For example, suppose that some lonely explorer discovers a new species, and she names the species in a letter she sends back home. We might want to say that when she wrote down the name in the letter, she was expressing a concept, even though at the time of her writing no one else had any awareness of the species she was naming, so no one else possessed the concept she was expressing.

But even though this concept might in fact be possessed by only one person, and so be private in one sense, there is another important sense in which that concept is *not* private—in which it *is* social—at least on my theory. That is because the concept will not always be possessed only by this one person, at least if she succeeds in communicating her discovery. As soon as her letter is read and understood, the concept will be shared among multiple people. Even though the concept is not possessed by other people, it is practically accessible

to other people in a way that it would not be on other theories, including individualistic theories.

So there is clearly room on my theory for concepts that are private in the sense that they happen not to be shared among multiple people—including concepts that could be expressed in public language but simply have not been so expressed yet. And these concepts need not be importantly different in kind from concepts that are widely shared, since the same theoretical rules can apply to them. The question remains open, though, of whether and how there is room on my theory for concepts that are private in some deeper sense—concepts, perhaps, that cannot be expressed in public language. In the next section, then, I will answer this question.

5.5 Public and private concepts

Above in subsection 5.2.1, I claimed that social theorists of concepts face an explanatory dilemma. On the one hand, social theorists can attempt to reduce the social entities their theories posit to more basic, individualistic entities, in which case they run the risk of an infinite explanatory regress. And on the other hand, social theorists can claim that at least some of the social entities they posit are explanatorily primitive, in which case they run the risk of metaphysical spookiness.

Now that I have presented the heart of my theory and demonstrated some of its consequences and applications, I wish to present my favored approach to this dilemma. However, before I do I should repeat my earlier disclaimer: this approach is not universally favored among social theorists, and it is not the only approach that is compatible with my theory as I have presented it so far. This section, then, will be even more of a speculative digression than the rest of this chapter.

5.5.1 There is danger of an explanatory regress

I would like to start working up to my solution by bringing the problem into somewhat clearer focus. First, I will discuss the precise theoretical constraint imposed by the danger of an explanatory regress, in particular as it applies to my theory.

In my original presentation of the dilemma, with a toy version of a social theory of concepts, the danger of explanatory regress arose because the theory sought to account for concept sharing as an intentional—and so conceptual—matter. To very briefly recap: If standing in the concept sharing relation to another thinker depends on having an intention that involves the concept CONCEPT, then explaining how a thinker possesses the (widely shared) concept CONCEPT—or how a thinker shares any other complex concept—will quickly enter a problematic regress. So the social theorist's account of the concept sharing relation should not depend on thinkers already possessing any widely shared complex concepts.

5.5.1.1 A metasemantic digression

Before I continue this discussion, I would like to take a brief detour through metasemantics. A somewhat similar problem arose in my discussion of Peacocke's attempt to account for the publicity of concepts in light of the fact that many people fall short of mastering many concepts: If being able to use the same concept as some other thinker depends on being able to use a word that has the same semantic value as a public-language word the other person is able to use, then explaining how an ignorant person is able to use a word with the same semantic value as an expert's word will be very difficult.

Since these lines of reasoning do not only apply to the toy social theory and Peacocke's theory, respectively, one upshot is that any social theorist's account of the concept sharing relation should not depend on thinkers already being able to use any term with the same semantic value as any particular public-language term. But our ways of identifying concepts according to their semantic values—such as naming them with a public-language word, as in "the concept CAT;" or using public language to describe what they are concepts 'of,' as in "her concept of cats"—identify concepts according to the semantic values they share with public-language words. So really, to be sure to avoid a problematic regress, any social theorist's account of the concept sharing relation should not depend on conceptual semantics. The concept sharing relation should, in effect, be presemantic. And since my theory identifies the concept sharing relation with the same term relation, one upshot for my theory is that the same term relation should be presemantic.

Now, in claiming that the same term relation should be presemantic, I might seem to be claiming that the relation and the semantics of the terms it relates are entirely independent of one another. This would put me in disagreement with Fine (2007), who as part of his "semantic relationism" theory proposes a relation that holds between terms and that functions importantly similarly to the same term relation I am proposing. As the name of the theory might suggest, Fine holds that this relation is an importantly semantic relation, in that standing in the relation is a semantic feature of a term.[24] Now, I do disagree with Fine on this point, because I hold that there is a real distinction between two term uses that are uses of a single term and two term uses that are uses of different terms, and I hold that any attempt to settle the semantic features of a term used in a given instance will depend

[24]For example, in discussing the phenomenon of "same-as representation," Fine claims that "the phenomenon is indeed semantic. When a piece of discourse represents an object as the same, then this is a semantic feature of the expressions by which reference to the object is made." (2007, 40)

on first coming to grips with this distinction.[25] But this metaphysical disagreement is not as important as it might seem, since like Fine, I hold that this relation that lies at the heart of my theory is very closely related to semantics—and indeed does depend on semantics in a limited way, as will become clear.

5.5.1.2 The constraint in application

Returning to the discussion, then: In developing my theory above, I claimed that the same term relation is a component of a rule in a game we are playing, viz. the game in which we treat ourselves and others as social, thinking, linguistic creatures. Our playing this game according to this rule will likely depend at least in part on our intentions and other thoughts, at the very least since even though it is not written down or otherwise publicly codified, the rule seems to extend to cases we never actually encounter when playing the game.[26]

The way the constraint applies to my theory, then, is that if I want to account for the same term relation as a component of a rule in this game we are playing, I will somehow need to account for our playing a game with these rules without depending on our already possessing any widely shared complex concepts. But since playing such a game is most naturally understood as an intentional matter, it is not immediately obvious how such an account can be given—at least if that account is not to involve explanatorily primitive social entities, such as the entities Brandom posits in his social theory.

This theoretical constraint should be familiar from earlier in this chapter, but as my mention of Peacocke above suggests, it should also be somewhat familiar from my discussion of the limitations of individualistic theories in earlier chapters—at multiple points in that discussion, individualistic theories were blocked from giving certain explanations of widespread concept sharing because those explanations themselves depended on widespread concept sharing.

As a way of building up to my solution to this problem, I now want to highlight a limit to the line of reasoning that led up to the identification of this constraint. This limit was not previously relevant because up until this point I have been focused on publicity as a desider-

[25]Fine seems to hold that the order of explanation is actually the reverse of what I am suggesting:

> For what is it for the names in the relevant sense to be the same? As we have seen, it is not simply a matter of typographic identity or coreference. But then what else is required? Presumably that the names should represent the object *as* the same, which is just what we were after. (Fine, 2007, 41)

He seems to suggest that this semantic same-as representation is constitutive of being the same name or term in the relevant sense, as if being the same term is a complex phenomenon made up of semantic as well as syntactic components.

[26]Wittgenstein (1997) and many philosophers following him, including Kripke (1982), offer famous doubts about whether understanding someone as having thoughts—as having internal representations of some kind—is really important to or sufficient for accounting for the rules she seems to apply. But these doubts go beyond the limited scope of this project, in part because they should be just as troubling for the individualistic theorists I discuss as for a social theorist of concepts. Moving forward, I will take it for granted that rules of implicit games like this are grounded at least in part in intentions and other thoughts

atum, and so the conceptual explanations I have been discussing have aimed at accounting for concepts values that are importantly public—concepts that are only interesting insofar as they are widely shared.

This line of reasoning shows how accounting for someone being able to use widely shared concepts will be problematic if that account depends on that person already possessing those same widely shared concepts—or even possessing other widely shared concepts, in many cases. The upshot for my theory is that in accounting for the same term rule, I should not rely on anyone already possessing any widely shared complex concepts, since someone possessing those concepts would have to already be playing a game with that rule.

However, it does not follow from this line of reasoning that accounting for someone being able to use widely shared concepts will be problematic if that account depends on that person already possessing any concepts at all. Or to put it in the 'game' terms of my theory, it does not follow from this line of reasoning that in accounting for someone playing a game, I cannot rely on that person already possessing any concepts at all.

5.5.2 Private games are possible

So to take an extreme example, the constraint I identified above does not apply to a theory intended to explain how an individual could play some entirely private game, in which she is the only player and the only judge, for which she alone devised all the rules. If, say, an infant stranded alone on an island somehow manages to reach adulthood without any human contact, and to keep herself amused she devises and plays some kind of game, then a theorist could give an explanation of the rules of that game as importantly semantic entities without once mentioning the same term relation—mentioning how the relation holds or how it does not hold—and without violating these constraints. The theorist could do this using one of the individualistic theories of concepts I discussed in chapter 3, and as long as the theorist did not claim that in devising her game this stranded game player used any widely shared complex concept (or any other term that shares a semantic value with a complex term in the theorist's own language), I would have no complaint.[27]

An individualistic theory of concepts could provide a satisfactory account of an entirely private game because a private game need not involve any public concepts. And in the extreme example I described above, the game must be private because the thinker playing it is entirely isolated. But private games can also be played outside of such extreme situations. Individuals can devise and play their own private games in many areas of ordinary life, and at least in many cases, accounts of these games need not mention any public concepts.

One question that is relevant to this project, then, is whether the game we play in which we treat ourselves and others as social, thinking, linguistic creatures could be such a private game. If it could, then even though its rules are importantly conceptual (and so semantic)

[27]For example, if the theorist claimed that this person conceived of her activity as a game (employing the concept GAME), I would question how the stranded game player could have acquired this particular concept. But the stranded game player need not conceive of her activity as a game for it to be a game, especially if she conceives of it as something very like a game.

entities, I could account for them using the tools offered by an individualistic theory of concepts, thus avoiding any violation of the constraints I discussed above. Unfortunately for me, though, I cannot plausibly claim that this game is private in this way. After all, this is a game we are all playing, so it cannot be private even in the limited sense of not being played by more than one person.

However, even though that game is not itself private, I will account for it in a way that importantly involves private games. I claim that there is a game I play alone the rules of which are very similar to the rules of that public game—a game in which I treat myself and others as social, thinking, linguistic entities. This game is private in just the same way as the entirely private game I mentioned above. And I will ultimately account for the single public game we are all playing by showing how it is grounded in the separate private games we are each playing, including this private game I am playing. These public and private games, though, exist at different levels of explanation, in a way that should become clear.

5.5.2.1 The general shape of my account

Now that I have introduced these private games, I would like to give a high-level overview of how the private games fit into the larger account of the publicity of concepts I am offering. The account is developed in two stages, involving two distinct levels of explanation or theoretical perspectives. I will offer parts of this account in greater detail below, but due to its multi-level structure, I hope this overview will help avoid confusion going forward.

So, at the more basic explanatory level, from a theoretical perspective much like the one we get on individualistic theories of concepts: I am privately playing a game in which I treat myself and others as social, thinking, linguistic creatures; and other thinkers are each playing a very similar game. The explanations of our playing these private games involve no public entities and can be given from within an individualistic framework, so in giving them I need fear no problematic regress.

Each of these private games involves a rule very similar to the same term rule, grounding a relation very similar to the same term relation, and thanks to the widespread instantiation of this relation, from the perspective of someone playing one of these private games, many concepts are widely shared. Or at least, from such a perspective, there is a state of affairs very similar to many concepts being widely shared—at this stage, I do not yet have the explanatory resources to claim that a contingent thought involving public concepts, such as the thought that is expressed by the public-language sentence "many concepts are widely shared," will be true or false from the perspective of someone playing one of these private games.

But instead of worrying about thoughts expressed using public-language sentences, for the next stage of the account, I will adopt my own perspective as someone playing one of these private games—adopt the folk theory my private game embodies—and worry about my private thoughts as they would be expressed using sentences in my own idiolect, such as the thoughts that ground my private game. To reiterate: from the first theoretical perspective I discussed above, these are thoughts built out of concepts that can be accounted for using

the tools of an individualistic theory. I will discuss their relationship to public concepts in more detail below, but for now I will just say that these concepts are importantly distinct from the public concepts that I have been discussing throughout this work as a theoretical goal.

So from my perspective as the player of my private game, the private thought I would express with the idiolectal sentence "many concepts are widely shared" is true. Continuing from that perspective and in my idiolect, then: Thanks to the way the same term relation is instantiated, the concepts that are widely shared include the concepts I possess that make up the thoughts that ground the rules of this game I seem to be playing. Indeed, these concepts are shared by other thinkers who are playing games—games grounded in thoughts made up of these very concepts. Since the thoughts, on close inspection, turn out to be made up of the same concepts in the same ways, they are the same thoughts. And since the games, on close inspection, turn out to be grounded in the same thoughts (and to be played in relevantly similar circumstances), they are the same game. Thus this game I am playing is actually a game we are all playing.

So at the first stage, from adoption and application of an individualistic theory of concepts, I get private concepts that make up private thoughts that ground a private game I am playing. And at the second stage, from adoption of the perspective of the player of that game and application of its rules, I get public concepts—which are distinct from the private entities I got at the first stage—that make up public thoughts that ground a public game we are all playing.

5.5.2.2 The privacy of my game

So before I can account for the public game we are all playing, I must first account for the private game I am playing. As I suggested above, if I do so using the tools of an individualistic theory of concepts, this should be unproblematic as long as I do not involve any public concepts in my account. This is a reasonable goal because I am the only one playing this game, and so I do not need to share with any other people the intentions and other thoughts that ground the game—just as it was with the entirely isolated individual I discussed at the beginning of this section, who was the only one playing her game.

There are, of course, some important differences between that imagined entirely private game and the game I play in which I treat myself and others as social, thinking, linguistic creatures; and these differences might suggest that the latter game could not be private in the relevant way. Before I move on, then, I will briefly discuss two of these differences.

One difference is that in the game I play, I assign significance to the actions of others as well as to my own actions, unlike the entirely private game played by an individual stranded on a deserted island. But a little reflection shows that this difference is not relevant here, because in assigning significance to the actions of others I need not use any public concepts—as long as "assigning significance to the actions of others" is not read so strictly that I must use the public concept ACTION, rather than some very similar concept, to do so. This

difference does highlight a limited sense in which the game I play is not private, but that sense of "private" is not relevant here.

The other difference builds on the first and is more relevant: In the game I play, I not only assign significance to the actions of others; I assign significance to the actions of others in just the same way that I assign significance to my own actions, treating all of us as players. Indeed, it might seem strange for me to posit a private game that I alone am playing, because it might seem that the only game is the game we are all playing in which we treat each other and ourselves as social, thinking, linguistic creatures. After all, many philosophers, linguists, and others seem to devote significant effort to studying such a game, which makes less sense as an endeavor if each is really only studying some private game that she and no one else is playing.

But as I suggested above, if this is a game that we are all playing intentionally, then it will be very difficult or impossible to account for the rules of the game without first accounting for how the same term relation is instantiated. This is because the game is informal, with its rules only implicit in its play and in the minds of those playing it intentionally, so if we are all playing the game intentionally, the rules will depend in some way on intentions and other thoughts shared by all or many of us players—shared intentions that involve shared concepts and so depend on the same term relation. And, of course, it will be very difficult or impossible to account for the same term relation before accounting for the rules of the game if, as I claim, the same term relation is nothing more than a component of a rule in the game.

I will not, then, claim that we are all intentionally playing the private game I have been discussing in which I treat myself and others as social, thinking, linguistic creatures. But neither will I deny that when I play the game I treat all of us as players. Instead, I claim that I am playing my game intentionally but others are playing it unintentionally—just as I am unintentionally playing the very similar games of others, each of whom is playing her own game intentionally. We are all intentionally playing our own private games, and we all seem to be playing the same game together because the games are so similar and because each involves all of us as players. Additionally, there is an important way in which we truly are all playing a single public game, which I will discuss in the next subsubsection—but because the publicity of this single game depends upon the publicity of concepts, this particular game cannot enter into my account here.

It might seem strange to talk about unintentional game playing, because prototypical game playing is intentional. For example, someone moving playing cards around a tabletop who merely happens to move them in accordance with the rules of solitaire might not be said to truly be playing solitaire, but someone moving the cards in the same way because she intends to play solitaire is playing solitaire. But a game can involve players who are not playing—or at least who are not playing that game in particular—intentionally. For an example of unintentional game playing that is perhaps a bit extreme, while watching students in a lecture class I could devise a game in which 'players' simply score a point for each minute they look in the direction of the lecturer. This example and others suggest that who counts as a player in a game is determined not by intentions—or at least not only by

intentions—but by the game itself, since many games explicitly or implicitly include in their rules some specification of who counts as a player. And as my own play shows, in the game I play in which I treat myself and others as social, thinking, linguistic creatures, all it takes for me to treat something as playing the game is for that thing to exhibit (or seem likely to exhibit) behavior that very roughly approximates cognitive or linguistic behavior.

5.5.3 Within the private game, concepts are public

So the picture I am offering is one in which the publicity of concepts depends on the widespread instantiation of the same term relation. This relation is merely a component of the game I play in which I treat myself and others as social, thinking, linguistic creatures. The rules of this game are not explicitly codified in some rulebook but are instead determined by my intentions and other thoughts, just as the rules of the very similar games that others are playing are determined by their intentions and other thoughts. These thoughts involve many complex concepts, but because it is not necessary for these concepts to be public (at least in this part of the picture), they can be accounted for using the tools of an individualistic theory of concepts, such as Fodor's theory or Peacocke's theory.

But among the rules of this private game is a rule very similar to the same term rule I discussed above—a rule to which all my above analysis of the same term rule would apply, if that analysis were made in my private idiolect, expressing my private thoughts that can be accounted for entirely individualistically, rather than in English. And so from my perspective as someone playing a game with that rule, concepts are public, because so many concepts are linked across thinkers by the same term relation that is grounded in that rule.

The publicity of concepts thus depends on concepts that are private, both in the sense that they are very likely not widely shared and in the sense that they arise out of an individualistic framework. And the publicity thus accounted for exists only as a feature of this private game I am playing, though the very similar games others are playing will have very similar features.

In an earlier section of this chapter, I described the theory I am offering as a folk psychological theory. I claimed that the theory was aimed at taking the rough and implicit rules of a game we are all playing and making those rules more explicit while smoothing out their inconsistencies and other problems. In an effort to avoid a problematic regress lurking in the vicinity of these rules—in an effort to work within the constraints I discussed above in this subsection, as well as elsewhere in this work—I have developed a theory that so far falls somewhat short of my stated aims, because the theory has not thus explicated a game we are all playing, instead only focusing on a game I alone am playing.

But it seems to me that this should still count as a folk psychological theory, because my game rule analysis was imprecise enough—and each of us is playing a game similar enough to the game I am playing—that the theory can cover games we are all playing. Additionally, as I mentioned above and will discuss in more detail in the next subsection, in part thanks to accounting for these private games, the theory will in the end be able to account for a single public game we are all playing—though like the publicity of concepts in general, the

publicity of this game will in some sense only exist as a feature of my private game, with others' very similar games having merely very similar features.

5.5.4 Private and public concepts are closely related but distinct

At this point I need to say more about the relationship between private and public concepts. As I stated above, the publicity of concepts depends on private concepts, because the publicity of concepts is a feature of a game I am playing in virtue of some intentions I have that involve private concepts. These concepts are private because they are sustaining the only game in which publicity is achievable. As I see it, in accounting for the publicity of any complex concept, a theorist must first account for this game, and in accounting for this game a theorist must account for these concepts. So these concepts exist outside the game in some sense, and they must have their semantic values and other conceptual features independently of the game.

These game-sustaining concepts are surely not the only concepts that can have some kind of existence outside of the game, appropriately analyzed using the tools of an individualistic theory of concepts. For instance, as I described in a case above, an explorer might have discovered a new species but not yet have expressed the name of that species publicly, so that an individualistic theory of concepts might be able to give a full and accurate analysis of the explorer's yet-to-be-shared concept. And at least in some important sense of "thought," there are many thoughts that are subpersonal and not even potentially communicated, and these thoughts will likely involve many concepts that are private in an even stronger way, again the appropriate objects of an individualistic theory. And as I described in another case above, it is surely possible (though surely extremely difficult) for an individual to live her entire life without mental contact with any other humans: there is no clear reason to treat the concepts of such a non-social individual as subject to the rules of the game in which I treat myself and others as social, thinking, linguistic creatures.

But there is an important difference between at least some of the game-sustaining concepts and these concepts that are private because they are newly invented, subpersonal, or possessed by an isolated individual. These latter private concepts are private at every level of analysis: not only do they come out not being widely shared on an individualistic theory, but also they should come out not being widely shared once the instantiation of the same term relation has been settled. These concepts as possessed by those individuals are not connected in the right way to any concepts as possessed by other individuals to be related by the same term relation, because the relevant causal connections or intentions do not exist. One easy way to see that the same term relation is not instantiated is that (intuitively) these concepts are not expressed in public language, at least yet.

On the other hand, some of the game-sustaining private concepts are at least very closely related to concepts that are expressed in public language. For example, as I described the same term rule above, it requires some kind of similarity in use, some kind of intention, and some kind of causal connection, or at least something very like these things. It would be very strange for me to claim that my thoughts that sustain this rule involve concepts that are

entirely distinct from the concepts I express using "similar," "intention," and "cause"—for one thing, a mind that has the public concepts SIMILAR, INTENTION, and CAUSE in addition to a trio of nearly indistinguishable private counterparts of those concepts seems implausibly uneconomical. How, then, are the private game-sustaining concepts related to public concepts like SIMILAR, INTENTION, and CAUSE? If they are distinct, why are they distinct? If they are the same, how can they be both public and private?

In a way, the answer to this question goes back to my discussion of kinds of theories of concepts in the last chapter. There, I made the point that there can be multiple theories each of which is importantly different from the others, each of which is unobjectionable, and each of which is appropriately called a theory 'of concepts.' And even though there might be a great deal of overlap in the subject matters of two theories, and even though the two theories might have very different things to say about their common subject matter, it might not be appropriate to view the two theories as contradicting one another. In that discussion, I was most interested in comparing what might be called 'psychological' theories to what might be called 'philosophical' theories of concepts. In this discussion, I am comparing different philosophical theories—or, more accurately, different parts of a single philosophical theory—but the point is still relevant.

The relationship between the private game-sustaining concepts and public concepts like SIMILAR, INTENTION, and CAUSE, then, is much like the relationship between a concept a psychologist might study and the corresponding concept a philosopher might study. They are not entirely distinct entities, since either they share a single realizer (if they are abstract) or they are one and the same token that falls under two types (if they are concrete) or one realizes the other (if one is concrete and the other is abstract). Furthermore, they are both theoretical entities, and their theories have much in common. But it can nevertheless be appropriate both to ascribe a certain property to one entity while working within one theory—such as ascribing privacy while working within an individualistic theory—and to ascribe an apparently incompatible property to the other entity while working within the other theory—such as ascribing publicity while playing the game in which I treat myself and others as social, thinking, linguistic creatures. So there is nothing contradictory or mysterious here, only a theory that must involve concepts in different ways at different stages of explanation because in order to understand ourselves as social, thinking, linguistic creatures, we must first understand ourselves as thinking creatures capable of devising and playing games.

As I discussed above, then, there are two theoretical perspectives from which concepts should be discussed in this theory. First, there is the explanatorily prior perspective from which our concepts are almost all private. It is appropriate to adopt this perspective when treating concepts that are not-yet-expressed, subpersonal, or otherwise outside of the game of treating oneself and others as social, thinking, linguistic creatures. An individualistic theory of concepts can be adopted for this perspective, such as Fodor's theory or Peacocke's theory.

Second, there is the perspective of someone playing the game of treating herself and others as social, thinking, linguistic creatures. Although the metaphysical underpinnings of

this perspective might include the other, individualistic perspective, this is the perspective we operate from in ordinary life—or at least this is a more explicit, consistent version of the perspective we operate from in ordinary life. The publicity of concepts is only realistically achievable from this perspective, as are many other components of our ordinary picture of ourselves. This is the perspective that is special to a social theory of concepts and that allows a social theory of concepts, unlike an individualistic theory of concepts operating on its own, to achieve the theoretical goals so common in philosophy that I discussed in the last chapter.

Before I conclude this subsection, I want to make good on a pledge I made above: When I discuss the same term relation—or when I discuss anything publicly—I am playing the game in which I treat myself and others as social, thinking, linguistic creatures. And from my perspective as I play that game, the concepts that make up my thoughts that sustain the same term rule and the other rules of that game might count as public, thanks to the instantiation of the same term relation. So if, as I suspect, the same term relation is instantiated in the right way, and we thus all share the thoughts that sustain the rules of the games we all play, then from within the game I am playing, we are all intentionally playing the same game. And similarly, from within the private game each of us is playing, we are all intentionally playing the same game—in a way, the private games pull themselves up to publicity by their bootstraps. It was, then, appropriate that I spoke of 'the game we play' above, even if it took until this point for me to have the resources to explain how it was appropriate.

5.5.5 CAT is a product of this public game

The explanatory structure of the view I am offering in this section is strange enough that the view is likely not yet entirely clear. So in order to help clarify the view, I will end this section by working through a somewhat concrete example. I will work through the example in what I think is the most natural order of analysis, beginning at the higher, public level at which we already find ourselves and then proceeding to the lower, private level. This is in contrast to the order I used above, where I began at the lower level and then moved on to the higher level.

As I have elsewhere in this work, I will use the example of the concept CAT and word "cat." However, unlike elsewhere in this work, here I will only be focusing on the entirely ordinary case of a competent English speaker who—though by no means a master of all things cat-related—suffers from no special confusions when it comes to cats.

So let us imagine that I am speaking with someone who seems to be a competent speaker of English and who at multiple points in the conversation makes the sound "cat" in some of the ways that ordinary English speakers do. In this situation, I will (at least implicitly) take this person to share the concept CAT with me and many other people, and I will continue the conversation accordingly, for example answering "I have two cats" when this person asks "How many pets do you have?" and forming the belief that this person has no cats when this person says "I have no cats." But why is it appropriate that I do these things?

I do these things because I am playing the game in which I treat myself and others as social, thinking, linguistic creatures. As someone playing that game, I follow the rules of the game, including the same term rule which dictates that in situations like this—in situations in which two players use terms with the right similarities in their use patterns, with the right intentions regarding publicity, and with the right causal connections—players count as sharing a term. Although I do not have the strongest evidence possible on the matter, because this is an ordinary situation I take this person and myself to meet the requirements of the same term rule with respect to "cat." I also take this person and myself to meet the requirements with respect to CAT, because I assume that we are each expressing a single concept with our uses of "cat" and so I extend the application of the same term rule to our concepts as well as to the words we are speaking. And other rules of the game dictate what behavior is appropriate in light of the sharing of this term—for example, that I should make inferences involving the term when the other person expresses thoughts involving the term.

So from my perspective as someone playing this game, the two of us count as sharing CAT. For similar reasons, from my perspective as someone playing this game, the two of us count as sharing many other concepts. And in part because of this, as well as because of other similarities in our thoughts and behavior, from my perspective as someone playing this game, the other person is playing the game too—they are following the same rules as I am. The game is thus a public game, at least from the perspective of someone playing it. But this is a perspective we are interested in when we do philosophy, because when we do philosophy we are often primarily interested in ourselves and others as social, thinking, linguistic creatures. So then, much of our philosophical analysis will be done from this perspective, including when the aim of that analysis is simply to make explicit the rules of the game, as my aim has been in this chapter. But how is it possible for us to have this perspective?

This is where the explanation must move to the lower, private level. At the higher level, the explanation takes my perspective as someone playing a game, and from that perspective, it turns out that many other people are playing the game too. But in order to account for my playing the game in the first place, because game playing (at least of this kind) is an intentional matter, we must attribute intentional states to me. Because this part of the account must be explanatorily prior to the part in which I am already playing the game, in giving this part of the account we cannot avail ourselves of any of the objects that only exist within the game—objects including CAT, as well as TERM and other public concepts.

Because we cannot bring public concepts into our account at this level, there is little hope of attributing any public intentional states to me, and since it is only in virtue of having these intentional states that I am playing the game I am playing, there is no reason to expect me to be playing a public game. Instead, from this perspective outside the game in which we treat ourselves and others as social, thinking, linguistic creatures, all we can account for are private concepts and private intentional states. But we can use these objects to do many things—to predict and explain individual behavior, for instance, and to account for private games. And once we account for these private games, we have given an objective metaphysical foundation to the account of the publicity of concepts we can only give from

the perspective of someone playing one of these games.

5.6 Conclusion

That, then, is my social theory of concepts in rough outline and rough application. Although it is far from being as thoroughly developed as some of the individualistic theories I have discussed, I hope the differences between it and those individualistic theories are clear enough that a useful comparison can now be made.

As I argued in chapters 2, 3, and 4, individualistic theories of concepts have a serious problem when it comes to the publicity of concepts: publicity is an explicit or implicit desideratum for many of these theories, but they are incompatible with it. Instead of a picture on which many concepts are truly shared among many people, these theories at best give us a picture on which many people possess many concepts that are each merely very similar to concepts possessed by others, with true concept sharing only happening in the case of simple concepts like ONE (and perhaps like ELECTRON) and very rarely by chance in the case of complex concepts like CAT or KNOWLEDGE.

On the other hand, social theories can give us a picture on which many concepts, including many complex concepts, are truly shared among many people. In the terms I introduced in chapter 1, social theories are compatible with 'minimal' publicity. But additionally, social theories allow for what the 'robust' publicity that I described in that chapter, since they give us a picture on which concepts are shared transitively, among large groups, easily, in spite of other differences, and thanks to social context. Social theories are thus able to respect our ordinary picture of ourselves as social, thinking creatures, and they can achieve this precisely by respecting this picture: my theory accounts for publicity simply by taking seriously the ways I treat people as sharing concepts.

And the costs associated with a social theory, or at least with the social theory I have outlined here, are not serious when weighed against the benefit of being able to account for publicity. True, my theory gives us a picture on which concept sharing is in some cases somewhat complicated or not fully determinate. But this is unavoidable and even desirable, at least if the theory is meant to do justice to our ordinary picture of ourselves. And true, my theory accounts for the publicity of concepts in part by positing a private game that depends for its existence on private concepts. But this is in part due to my desire to give a metaphysically reductive account of social phenomena in terms of individualistic phenomena, so this multilayered approach should seem reasonable to anyone else who holds that our ability to think is more basic than our picture of ourselves as social, thinking creatures. Anyone who denies this, instead viewing all thought as necessarily social, would likely be more attracted to other social theories of concepts[28] anyway, since this view is more compatible with social than with individualistic theories.

[28]For example, as I already mentioned, Brandom (1994) offers a social theory that does not have any role for private concepts.

And when desiderata other than publicity are considered—for example, the desideratum of assigning intuitive semantic values to concepts—there is no reason to prefer either social or individualistic theories. As I developed my theory above, it remains largely neutral on issues such as how, if at all, semantic values should be assigned to concepts. These issues are orthogonal to the main point of contention between social and individualistic theories as families of theories of concepts, which is whether and how publicity is to be accounted for. Individualistic theorists like Fodor and Peacocke, for example, could adopt a social approach much like the one I presented here without greatly changing the ways they assign semantic values to concepts.

So on balance, the benefits associated with social theories as a family of theories of concepts far outweigh the costs. Moving forward, then, philosophers concerned with the publicity of concepts should adopt social theories of their own, hopefully developing them further than I developed mine in this work.

Bibliography

Armstrong, D. M. (2010). *Sketch for a Systematic Metaphysics*. Oxford University Press.

Block, N. (1986). Advertisement for a semantics for psychology. In P. A. French, T. E. Uehling Jr., H. K. W., editor, *Midwest Studies in Philosophy X: Studies in the Philosophy of Mind*. Minnesota: University of Minnesota Press.

Boghossian, P. (1994). Inferential role semantics and the analytic/synthetic distinction. *Philosophical Studies*, 73:109–122.

Brandom, R. (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4(1):73–121.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

Dretske, F. (1986). Misrepresentation. In Bogdan, R., editor, *Belief: Form, Content, and Function*. Oxford University Press.

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

Fine, K. (2007). *Semantic Relationism*. Blackwell.

Fodor, J. (1990). *A Theory of Content and Other Essays*. MIT Press.

Fodor, J. (1995). *The Elm and the Expert: Mentalese and its Semantics*. MIT Press.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press.

Fodor, J. (1999). Information and representation. In Margolis, E. and Laurence, S., editors, *Concepts*, chapter 22, pages 513–524. MIT Press.

Fodor, J. (2004). Having concepts: a brief refutation of the twentieth century. *Mind & Language*, 19(1):29–47.

Fodor, J. and LePore, E. (1992). *Holism: A Shopper's Guide*. Blackwell.

Frege, G. (1956). The thought: A logical inquiry. *Mind*, 65:289–311.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66:377–388.

Grice, H. P. (1968). Utterer's meaning, sentence-meaning, and word-meaning. *Foundations of Language*, 4:225–242.

Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review*, 78:147–177.

Harman, G. (1973). *Thought*. Princeton University Press.

Harman, G. (1982). Conceptual role semantics. *Notre Dame Journal of Formal Logic*, 23:242–256.

Kaplan, D. (1990). Words. *Proceedings of the Aristotelian Society, Supplementary Volumes*, pages 93–119.

Kripke, S. (1979). A puzzle about belief. In Margalit, A., editor, *Meaning and Use*, pages 239–83. Reidel.

Kripke, S. (1982). *Wittgenstein on rules and private language*. Harvard University Press.

Lackey, J. (1999). Testimonial knowledge and transmission. *The Philosophical Quarterly*, 49(197):471–490.

Laurence, S. and Margolis, E. (2002). Radical concept nativism. *Cognition*, 86:25–55.

Lewis, D. (1975). Languages and language. In Gunderson, K., editor, *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4):343–377.

Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell.

Loar, B. (1981). *Mind and Meaning*. Cambridge University Press.

MacFarlane, J. (2007). Relativism and disagreement. *Philosophical Studies*, 132:17–31.

Madagáin, C. Ó. (2011). *The Public Dimension of Meaning*. PhD thesis, University of Toronto.

Peacocke, C. (1995). *A Study of Concepts*. MIT Press.

Putnam, H. (1973). Meaning and reference. *Journal of Philosophy*, 70(19):699–711.

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7:131–193.

Putnam, H. (1980). Models and reality. *The Journal of Symbolic Logic*, 45:464–82.

Schiffer, S. (1972). *Meaning*. Clarendon Press.

Schiffer, S. (1987). *Remnants of Meaning*. MIT Press.

Sider, T. (2011). *Writing the Book of the World*. Oxford University Press.

Smith, N. J. J. (2008). *Vagueness and Degrees of Truth*. Oxford University Press.

Sorensen, R. (2001). *Vagueness and Contradiction*. Oxford University Press.

Stich, S. (1983). *From Folk Psychology to Cognitive Science*. MIT Press.

Travis, C. (1985). On what is strictly speaking true. *Canadian Journal of Philosophy*, 15:187–229.

Unger, P. (1980). The problem of the many. *Midwest Studies in Philosophy*, 5:411–67.

Williamson, T. (1994). *Vagueness*. Routledge.

Wittgenstein, L. (1997). *Philosophical Investigations*. Blackwell. Translated by G. E. M. Anscombe.