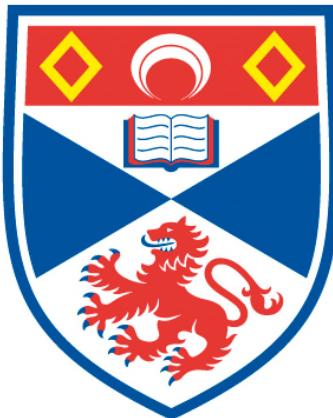


CONCEPTS IN CONTEXT

Andrea Onofri

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**



2013

**Full metadata for this item is available in
St Andrews Research Repository
at:**

<http://research-repository.st-andrews.ac.uk/>

**Please use this identifier to cite or link to this item:
<http://hdl.handle.net/10023/3462>**

This item is protected by original copyright

CONCEPTS IN CONTEXT

Andrea Onofri

Submitted for the degree of Doctor in Philosophy

University of St. Andrews, August 2012

Candidate's declarations:

I, Andrea Onofri, hereby certify that this thesis, which is approximately 87.723 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2008 and as a candidate for the degree of Doctor in Philosophy in September 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2012.

Date: 17/8/2012

Signature of candidate: Andrea Onofri

Supervisor's declaration:

I, Herman Cappelen, hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor in Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date:

Signature of supervisor:

Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to all of printed copy but embargo of all of electronic publication of thesis for a period of 2 years on the following ground:

Publication would preclude future publication.

Date: 17/8/2012 Signature of candidate: Andrea Onofri

Date: Signature of supervisor:

Abstract

My thesis tackles two related problems that have taken center stage in the recent literature on concepts:

- What are the *individuation conditions* of concepts? Under what conditions is a concept C_1 the same concept as a concept C_2 ?
- What are the *possession conditions* of concepts? What conditions must be satisfied for a thinker to have a concept C ?

I will develop a *pluralist* and *contextualist* theory of concept individuation and possession: different concepts have different individuation and possession conditions, and contextual factors play a crucial role in determining what concepts we attribute to other subjects when we ascribe propositional attitudes to them.

In **chapters 1-3**, I defend a *contextualist, non-Millian* theory of propositional attitude ascriptions. Then, I suggest contextualist theories of ascriptions can be applied to the problem of concept individuation/possession. In particular, I use contextualism to provide a new, more effective argument for Fodor’s “publicity principle”, according to which concepts must be *shared* in order for interpersonally applicable *psychological generalizations* to be possible.

Publicity has important implications: in particular, it is inconsistent with existing versions of *holism*, on which concepts *cannot* be shared by ordinary thinkers. Nonetheless, in **chapters 4-5** I show how holism can still play an important role in our best theory of concepts. More specifically, I argue that the tradition of appealing to *modes of presentation* in order to give an account of “Frege cases” is in fact *committed* to holism. To develop a version of holism that will give a successful account of Frege cases without violating publicity, I suggest we should adopt my pluralist-contextualist picture: on that picture, the concepts involved in a Frege case will be holistically individuated and not public, while other concepts will be more coarsely individuated and widely shared. In **chapter 6**, I will develop this view further by contrasting it with other *pluralist* theories (Weiskopf) and with rival theories of concepts, such as the *localist* views defended by Peacocke, Rey and Jackson.

Acknowledgments

I am delighted to have the chance to finally thank all the people who made this dissertation possible. As usual, they are way too many to mention here, and I apologize in advance with those whose names have been omitted.

One of the aspects of our profession that I find most rewarding is that we get the opportunity to collaborate with some truly extraordinary people. During these four years, I have been lucky enough to meet a great number of incredibly talented, passionate and kind philosophers. My conversations with them have shaped this dissertation, which is to a significant extent the product of their work¹.

My first thanks go to my previous supervisors, Berys Gaut and Simon Prosser, and to my current supervisors Herman Cappelen, Jessica Brown and Francois Recanati. Each of them has helped me in many different ways; I cannot hope to make full justice to their contribution, but I will do my best.

Berys Gaut has been my very first supervisor; had it not been for him, I would not have started my Ph.D in St. Andrews, nor would I have developed an interest in concepts. His support and encouragement, as well as his enlightening philosophical feedback, have been the first seed of this thesis. Simon Prosser has been my main supervisor for two years. His help has been absolutely crucial, especially in the early stages of my doctoral work: Simon believed in my philosophical potential from the very beginning, and I would never have found the strength to start a dissertation on concepts without his support. His patience and kindness in dealing with an often impatient and self-doubting student have been simply astonishing.

Herman Cappelen has been my main supervisor in the last year of the Ph.D, when most of this dissertation was written. His feedback shaped this work into its current state. All my supervision meetings with him have been illuminating: with a few words, Herman would explain to me how to structure an argument, a section, a chapter, and a whole host of problems that were worrying me would simply disappear. I am deeply grateful to him for believing in me and deciding to supervise me despite his several other commitments. Herman's desire to help his students and his effectiveness in helping them are unique: I have learned a great deal from him.

In theory, Jessica Brown should have been one of my second supervisors. In fact, she has been a lot more; we met on a regular basis to discuss various parts of the thesis, and her feedback has been invaluable in shaping many of the chapters. Jessica has a rare

¹ Because this dissertation has benefited from the help of so many people, I have preferred not to include any specific acknowledgments in the text, in order to make the discussion smoother and easier to follow for the reader.

ability to detect immediately where the problems with a certain philosophical position lie: many of the arguments in this thesis were greatly improved thanks to her suggestions and criticisms. I am also extremely grateful to Jessica for her constant support and encouragement, which have meant a lot for me in difficult times.

My other secondary supervisor, Francois Recanati, has also played a very important role in the development of this thesis. My meetings with him were invariably enjoyable and intellectually stimulating: I have rarely met someone so quick at getting the gist of an objection and coming up with an effective reply. Some of the core ideas developed in the central part of the thesis are the result of Francois' suggestions, and many of the chapters were written while having Francois' theory of mental files in mind.

Besides my supervisors, there are several other philosophers I would like to thank. Two of them deserve a special mention. Derek Ball has done a lot more than it was reasonable to expect in order to help me in my philosophical efforts; without his aid, I'm not sure I would have managed to complete this dissertation. He has read countless drafts and provided extensive feedback on all of the chapters. Derek has also been one of my main models in these years. I greatly admire his ability to be both clear and profound in his philosophical points, always paying attention to the details without losing sight of the bigger picture. I hope that, by working with him, I absorbed a tiny bit of his talent and of the intellectual honesty with which he uses it.

Another person to whom I owe a great debt of gratitude is Torfinn Huvenes. In retrospect, I could not have been luckier in getting Torfinn as my officemate when I first joined Arché. I have learned a lot from Torfinn; I will never stop being amazed by his knowledge, his intelligence and his passion for philosophy. My conversations with him were invaluable in developing this thesis: many of my main ideas derived more or less directly from his suggestions. Whenever I asked for his assistance, he was always there to help, giving me advice about what to read on a certain topic, how to patch an argument or how to state a thesis. It's great to finally have the chance to thank him for being such a good friend and philosophical mentor at the same time.

Next, I would like to thank all the other people who made my Ph.D at St. Andrews such an amazing experience. First, thanks to all the participants in the Arché project on "Philosophical Methodology". It has been a privilege to collaborate with the project, working with a group of amazing philosophers on such an interesting topic. Next, I would like to thank all the professors, postdocs and PhD students at the Arché Research Centre, which has been my home over the past three years. My intellectual experience at Arché has been extraordinarily intense and rewarding, and I will never be grateful enough to Jessica Brown and Herman Cappelen for giving me the opportunity to join the centre. First, let me thank all my fellow Ph.D students and friends at Arché and St. Andrews, particularly Daniele Sgaravatti, Steven Hall, Nick Hughes, Joshua Thorpe, Laura Porro, Laura Celani, Margot Strohminger, Anders Schoubye, Andreas Stokke, Thomas Hodgson, Mark Bowker, Noah Friedman-Biglin, Michael De, Andri Hjálmarsson, Daniele Labriola, Ines De Asis, Simon Fokt, André Grahle, Christopher

Woerner. Special thanks go to Dirk Kindermann, whose help over the past 12 months has been simply invaluable; and to Julia Langkau, whose support and encouragement have meant a lot more than she thinks.

Next, let me thank our former postdocs Jonathan Ichikawa, Dilip Ninan and Yuri Cath, as well as our current lecturer Ephraim Glick, for their feedback on my work and for many, many great seminar sessions; and all our professorial fellows, particularly Brian Weatherson, who has helped me in a number of different ways during my doctoral studies. I would also like to thank all of our administrative staff, and particularly Shaun Darby and Lynn Hynd, for their patient help in all these years. Finally, thanks to the St. Andrews/Stirling Graduate Programme and the AHRC for their financial support.

I have been lucky enough to spend three months at the Rutgers philosophy department in New Jersey and two months at the Institut Jean Nicod in Paris. My experience in both places has been amazing, and there are several people I should thank for this. First, thanks to my supervisor at Rutgers, Jerry Fodor: for being so generous with his time and feedback; for being so supportive and encouraging; and for being such a great philosopher and writer. Next, thanks to the graduate communities at Rutgers and in Paris; unfortunately, in both of these places I made too many friends and interacted with too many good philosophers to mention them all here. Special thanks go to Ernest Sosa (Rutgers) and Pierre Jacob and Jerome Dokic (Institut Jean Nicod) for their very helpful feedback. Let me also thank Nicoletta Loccioni and all the members of the CPR project in Paris. Finally, a special mention goes to Michael Murez, one of the people with whom I most enjoy doing philosophy; my conversations with him have played a big role in my philosophical development during the last year and a half.

I would also like to thank my audiences in Paris, Princeton, Austin, St. Andrews, Milan, Padua, Reading, Boston for very useful feedback. In particular, thanks to David Gray, John Butterworth, Jeremy Goodman, Nat Hansen, Mark Harris, Travis Hobbs. I am especially grateful to Ray Buchanan, whose insightful and extensive feedback was crucial in improving chapter 1.

During the last four years, I have been able to rely on the help of an extraordinary group of Italian friends. Whenever I went back to Rome, my friends were always there waiting for me; thanks to them, it never really felt like I had left Italy. I owe a lot to all of them, and they probably don't realize what their friendship means for me. Grazie in particolare ad Angela, Chiara, Cristiano, Dario, Elena, Enrico, Federica, Flavio, Giangi, Giorgio, Ilaria, Lanfranco, Laura G., Laura M., Lavinia, Mirko, Pierfrancesco, Pierpaolo, Sisto, Stefania, Tommaso, Valeria.

My last and biggest thanks go to my family, whose loving support has been the most important ingredient in this dissertation: none of it would have been possible without their help. In particular, let me thank my grandparents Giovanni, Graziella, Luciano, Maria and my aunt Angiolina, who have all done so much for me in the past 27 years.

Let me also thank my brothers, Alessandro and Alberto, for making our house such a fun place to be in. And, finally, let me thank my parents Alfredo and Antonella. It would be impossible to explain how amazing they are in a few words. It will suffice, then, to say that they have managed to create a genuinely happy family; which I regard as one of the greatest accomplishments one can possibly hope for. This thesis is dedicated to them.

Table of contents

INTRODUCTION	16
CHAPTER 1: TWO VARIETIES OF MILLIANISM	29
1. INTRODUCTION	30
2. MILLIANISM: AN OVERVIEW	32
2.1 MILLIAN THEORIES AND THE SUBSTITUTION OBJECTION	32
2.2 A MILLIAN METAPHYSICS OF BELIEF	34
2.3 PRAGMATIC MILLIANISM	35
2.4 NON-PRAGMATIC MILLIANISM	36
3. NPM AND ENLIGHTENMENT: WHEN PSYCHOLOGY IS NOT ENOUGH	39
3.1 THE ASYMMETRY OBJECTION	39
3.2 FIRST REPLY: BITING THE BULLET	41
3.3 SECOND REPLY: ANTI-MILLIAN PRINCIPLES	43
4. NPM AND BEHAVIOR	49
5. SOME PROBLEMS WITH PRAGMATIC MILLIANISM	54
CHAPTER 2: PUBLICITY VS. HOLISM	59
1. INTRODUCTION	60
2. THE PUBLICITY PRINCIPLE	61
2.1 THE REPRESENTATIONAL THEORY OF THE MIND (RTM)	61
2.2 THE ARGUMENT FOR PUBLICITY	63
2.3 PUBLICITY AND HOLISM	67
3. TWO FAILED HOLISTIC RESPONSES	72
3.1 FUNDAMENTAL GENERALIZATIONS?	72
3.2 THE APPEAL TO SIMILARITY	79
4. BACK TO PUBLICITY	84
CHAPTER 3: PUBLICITY, CONTEXTUALIST-STYLE	90
1. INTRODUCTION	91
2. INTENTIONAL GENERALIZATIONS AND THE SEMANTICS OF REPORTS	92
2.1 TWO MORALS	92
2.2 CRIMMINS' THEORY OF ASRIPTIONS	97
3. A CONTEXTUALIST PICTURE OF PUBLICITY	102
4. HOLISM IN A CONTEXTUALIST FRAMEWORK	111
4.1 A REVISED VERSION OF HOLISM	111
4.2 SCHNEIDER'S APPEAL TO MILLIANISM	118
CHAPTER 4: PUBLICITY AND THE FREGEAN CONSTRAINT	124
1. INTRODUCTION	125
2. THE FREGEAN CONSTRAINT	129

2.1	FREGE CASES AND RATIONALITY	129
2.2	EXPLAINING RATIONALITY: THE FREGEAN CONSTRAINT	130
3.	INDIVIDUATION CRITERIA FOR CONCEPTS	136
3.1	SOME CRITERIA	136
3.2	SOME TERMINOLOGY: INDIVIDUATION, RESTRICTION, SATISFACTION	140
4.	A TENSION BETWEEN CONSTRAINTS	145
4.1	THE DOUBLE SUPERMAN CASE	145
4.2	SATISFYING (FC) Vs. RESPECTING (PUB)	146
4.3	A SOLUTION?	154
CHAPTER 5: SOLVING THE TENSION		158
1.	INTRODUCTION	159
2.	THE IMPLICATIONS OF (FC)	161
2.1	WHAT (FC) DEMANDS	161
2.2	A FINE-GRAINED CONDITION	166
3.	THE SOURCE OF THE PROBLEM	170
4.	CONCEPTS: A PLURALIST-CONTEXTUALIST PICTURE	175
4.1	TWO SOLUTIONS	175
4.2	A PLACE FOR HOLISM	181
4.3	A PICTURE OF CONCEPTS	189
CHAPTER 6: TWO DILEMMAS FOR LOCALISM		196
1.	INTRODUCTION	197
2.	THE DILEMMA FROM COGNITIVE PSYCHOLOGY	199
2.1	LOCALISM AND ITS VIRTUES	199
2.2	WHAT DOES PSYCHOLOGY TELL US ABOUT CONCEPTS?	201
2.3	WEISKOPF'S PLURALIST LOCALISM	203
3.	THE CORRECTNESS DILEMMA	210
3.1	A DUAL THEORY OF CONCEPTS	210
3.2	PLAUSIBILITY AND CORRECTNESS: A TRADE-OFF	212
4.	REFERENCE-FIXING CONCEPTIONS?	217
5.	CONCLUSION	226
APPENDIX A		230
APPENDIX B		231
REFERENCES		233

Introduction

Two problems loom large in the recent literature on concepts. The first one concerns their *individuation conditions*:

Individuation problem: What are the individuation conditions of concepts? Under what conditions is a concept C_1 *the same concept* as a concept C_2 ?

To give the reader just one very recent example, here is how Sainsbury and Tye present their “originalist” theory of concept individuation²:

Originalism answers the question: what are the necessary and sufficient conditions for the concept C_1 to be the same concept as the concept C_2 ? According to originalism, every concept has exactly one originating use, and every originating use of a concept is an originating use of just one concept. Hence we can offer the following necessary and sufficient condition for concepts to be the same:

(O) Concept C_1 = concept C_2 iff the originating use of C_1 = the originating use of C_2 .

In this picture, each use U of a concept is a use of the unique concept that lies at the origin of the R-linked chain of uses to which U belongs (Sainsbury and Tye 2011, p. 105).

A second problem concerns the *possession conditions* of concepts:

Possession problem: What are the possession conditions of concepts? What conditions must be satisfied for a thinker to *have* a concept C ?

Here, for instance, is how Peacocke describes the possession conditions of the concept CONJUNCTION³ on his “inferentialist” account (Peacocke 1992, p. 6. I slightly modified Peacocke’s notation to conform to my own notation for concepts):

CONJUNCTION is that concept C to possess which a thinker must find transitions that are instances of the following forms primitively compelling, and must do so because they are of these forms:

$$\begin{array}{ccc} p & \frac{p \ C \ q}{q} & \frac{p \ C \ q}{q} \\ q __ & p & q \\ p \ C \ q & & \end{array}$$

² Several other theories of concept individuation and possession will be discussed later (v. especially ch. 2 (sect. 2.3) and ch. 4 (sect. 3.1)).

³ I will refer to concepts by using expressions in small capitals.

(Simplifying: CONJUNCTION is that concept C to possess which a thinker must accept as valid all instances of the rules of conjunction introduction and elimination).

Let's say a bit more about our two questions. First, notice that there might not be a *single* answer to the interrogatives above. Different concepts might have different individuation conditions: it might be that, for a concept C to be the concept CONJUNCTION, C must satisfy certain necessary/sufficient conditions, where these conditions are *different* from those C must satisfy in order to be the concept BACHELOR or SUPERMAN. Even if concepts did have heterogeneous individuation conditions, however, the individuation problem would still arise. We would just have to state our question at the level of *specific* concepts, as in:

Individuation problem (CONJUNCTION): What are the individuation conditions of the concept CONJUNCTION? Under what conditions is a concept C the same concept as CONJUNCTION?

And so on for every other concept (BACHELOR, SUPERMAN...). The same holds for the possession problem. Different concepts might have different possession conditions: the conditions a thinker must satisfy in order to have the concept CONJUNCTION might be different from those he must satisfy to have BACHELOR or SUPERMAN. Again, this would not rid us of the possession problem, which will now arise at the level of specific concepts:

Possession problem (CONJUNCTION): What are the possession conditions of the concept CONJUNCTION? What conditions must be satisfied for a thinker to have CONJUNCTION?

(Indeed, I will argue later that *there isn't* a single answer to our two questions: concepts do have different individuation and possession conditions. V. especially ch. 5, sect. 4.3).

A second point to note is that our two problems seem to be closely related. As Fodor puts it:

It's a general truth that if you know what an x is, then you also know what it is to have an x. And ditto the other way around. This applies to concepts in particular: the question what they are and the question what it is to have them are logically linked; if you commit yourself on one, you are thereby committed, willy nilly, on the other. Suppose, for example, that your theory is that concepts are pumpkins. Very well then, it will have to be a part of your theory that having a concept is having a pumpkin. And, conversely: if your theory is that having a concept is having a pumpkin, then it will have to be a part of your theory that pumpkins are what concepts are (Fodor 1998, p. 2).

Fodor takes this to show that we can derive possession conditions for concepts from their individuation conditions:

Until quite recently (until this century, anyhow) practically everybody took it practically for granted that the explanation of concept possession should be parasitic on the explanation of concept individuation. First you say what it is for something to be the concept *x* - you give the concept's 'identity conditions' - and then having the concept *x* is just having whatever the concept *x* turns out to be (*ibid.*).

Conversely, Peacocke has argued that a theory of concept individuation should be grounded in a theory of concept possession:

Principle of Dependence: there can be nothing more to the nature of a concept than is determined by a correct account of the capacity of a thinker who has mastered the concept to have propositional attitudes to contents containing that concept (a correct account of "grasping the concept") (Peacocke 1992, p. 5).

The individuation conditions for a particular concept can then be straightforwardly derived from its possession conditions (*ibid.*, p. 6):

Accepting the Principle of Dependence opens up the possibility that we can simultaneously say in a single account what individuates a particular concept and also what it is to possess that concept. The general form that could be taken by such an account is this:

Simple Formulation: Concept *F* is that unique concept *C* to possess which a thinker must meet condition *A(C)*

(V. for instance his account of CONJUNCTION). Simplifying, we might describe the relation between individuation and possession conditions as follows. Suppose a concept *C₁* is "partially individuated" by its being *F*; that is, suppose it's true that, in order for a concept *C₂* to be the same concept as *C₁*, *C₂* must also have *F*. (For instance: if *o* is *C₁*'s origin, it follows from Sainsbury and Tye's account that a concept *C₂* will be the same as *C₁* only if *C₂* also has *o* as its origin). We can then infer that a subject will *have C₁* only if one of his concepts has *F*: if not, none of his concepts will be the same as *C₁* and our subject will lack the concept in question. (On Sainsbury and Tye's account, for instance, you have the concept BACHELOR only if one of your concepts has the same origin as BACHELOR).

Conversely: suppose condition *k* is part of the possession conditions of concept *C₁*, so that a thinker will have *C₁* only if he also satisfies *k*. (For instance: on Peacocke's account, one has the concept CONJUNCTION only if he is disposed to follow the rules of conjunction introduction/elimination). We can then infer that concept *C₁* is partially individuated by its possession condition *k*: for a concept *C₂* to be the same as *C₁*, *C₂* must also be such that, to have *C₂*, a thinker must meet *k*. (On Peacocke's account, for instance, a concept *C* cannot be the concept CONJUNCTION unless it's true that, to have *C*, a thinker must be disposed to follow conjunction introduction/elimination). Given the tight connection between individuation and possession conditions, then, we can expect a

theory of concept individuation to have direct consequences for the problem of concept possession, and vice versa: this will be a constant theme in our discussion.

A third important aspect of the individuation problem is that it should not be confused with a different issue, which we might call “the ontological problem”⁴. This amounts to the following question: what *kind* of entity is a concept? For instance: is it an abstract object? Or is it a concrete mental representation, e.g. a token symbol in our “Language of Thought”⁵? Clearly, the two problems are potentially related. An object’s individuation conditions might vary depending on its ontological domain: the conditions under which person P₁ is the same person as P₂ are presumably different from the conditions under which electron E₁ is the same electron as E₂. So an ontological theory *might* have consequences for a theory of concept individuation, and vice versa. Still, the individuation problem is clearly distinct from the ontological one, and it will be important to keep the distinction in mind throughout our discussion. (In fact, as we will see later, most theories of concept individuation can be formulated in terms that would be acceptable on *any* ontology of concepts. This will enable us to develop a theory of concept individuation/possession without taking a stance on whether concepts are abstract objects or mental representations).

Why has the problem of concept individuation and possession played such a major role in the recent literature on concepts? A first obvious reason is that a theory of concepts would seem lacking unless it could provide an account of their identity conditions and, relatedly, of the conditions that must be satisfied in order to have them. In turn, a theory of concepts can play a central role in a more general theory of *intentionality*. According to a major tradition in the philosophy of mind and language, concepts are the basic constituents of *thoughts*. In turn, thoughts are structured representations endowed with truth-conditions and acting as *relata* of propositional attitudes: the thought DOGS BARK is true just in case dogs bark, and to believe (rather than, say, desire or hope) that dogs bark is to stand in a certain relation to that thought⁶. On this approach, a theory of concepts will be an essential component of an account of propositional attitudes and intentional representations in general. For instance, to explain how people can have propositional attitudes about dogs we will first need to explain how they can have the concept DOG, since that will be a basic constituent of all their dog-directed thoughts.

But the ramifications of the problem of concept individuation/possession extend beyond the theory of intentionality, touching on several other areas of philosophy and

⁴ V. Laurence and Margolis (2007, 2011) for an overview. We will go back to this in ch. 2 (sects. 2.1-2.2).

⁵ Further options are available: for instance, one might identify concepts with abilities (Evans 1982, Millikan 2000). Our discussion will focus on the two options in the main text, although much of what I’ll say would also apply to other ontologies of concepts.

⁶ V. ch. 2 (sect. 2.1) for more details. This is by no means the only approach to the problem of intentionality: v. for instance Stalnaker (1984, 2008) for a radically different account.

on the methodology of philosophy itself. Suppose we are convinced by Peacocke's "inferentialist" account of CONJUNCTION, on which specific inferential dispositions are required for possession of the concept. We might then decide to extend the account to other concepts, e.g. the concept BACHELOR: for instance, we might hold that a subject doesn't have BACHELOR unless he is disposed to infer from $X \text{ IS A BACHELOR}$ to $X \text{ IS AN UNMARRIED MAN}$, and vice versa. (Simplifying: unless he believes that something is a bachelor *iff* it is an unmarried man). On this account, BACHELOR would then be individuated by its inferential connections to the concepts UNMARRIED and MAN: to put it a bit roughly, a concept C would not be the concept BACHELOR unless C 's owner was disposed to infer from $X \text{ IS } C$ to $X \text{ IS AN UNMARRIED MAN}$ (and vice versa).

The idea that certain inferential links are part of the individuation and possession conditions of concepts will be one of the main foci of our discussion. This venerable idea has played a crucial role in a number of philosophical debates. Here are just two of its possible consequences:

Epistemological consequences: Suppose the concept BACHELOR is indeed individuated by its connection to UNMARRIED and MAN; and suppose it's true that a thinker will not have BACHELOR unless he is disposed to infer from $X \text{ IS A BACHELOR}$ to $X \text{ IS AN UNMARRIED MAN}$ (and vice versa). Many have taken this to show that inferences like $X \text{ IS A BACHELOR} \rightarrow X \text{ IS UNMARRIED}$ are *analytic* (analytically valid), and that the corresponding thought EVERY BACHELOR IS UNMARRIED is *analytically true*⁷. In turn, the alleged analyticity of such inferences/thoughts has led many to ascribe a special *epistemological* status to them. Suppose the inference $X \text{ IS A BACHELOR} \rightarrow X \text{ IS UNMARRIED}$ is indeed a constitutive component of the concept's identity. Then, it seems we could come to know that inference to be valid (and the corresponding thought to be true) without having to investigate empirically whether bachelors are in fact unmarried; on the contrary, we could come to know that bachelors are unmarried in a purely *a priori* way, by simply reflecting on the "structure" of the concept BACHELOR.

Methodological consequences: Suppose philosophically interesting concepts such as TRUTH, KNOWLEDGE, JUSTICE are also individuated by inferential connections which are required for their possession (e.g. $X \text{ KNOWS THAT } P \rightarrow X \text{ BELIEVES THAT } P$). The above considerations would then show such inferences to be analytic and *a priori*. In turn, defenders of "conceptual analysis" have taken this to support their favored methodology⁸. Philosophical inquiry, they claim, is at least in part an inquiry into the structure of our concepts: our standard "armchair" methodology consists in using intuitions about whether concept C applies to case x and then reconstructing the

⁷ Of course, there are several ways to spell out the notion of analyticity (v. Williamson 2007, chs. 3-4 for a critical overview). The literature on the topic is huge: for a recent defense of the view, v. for instance Bealer (1998), Boghossian (2001, 2003), Jackson (1998), Peacocke (2007), Rey (2005).

⁸ Besides the authors quoted in fn. 7 *supra*, v. for instance Goldman and Pust (1998), Goldman (2007). For an overview, v. DePaul and Ramsey (1998).

inferential dispositions which must have produced that intuition. When conducted properly on a sufficient range of cases, this methodology will enable us to identify those fundamental inferential dispositions which are constitutive of our concepts and required for their possession. (For instance, Gettier cases show that the inference X HAS A JUSTIFIED TRUE BELIEF THAT P → X KNOWS THAT P is *not* part of our concept KNOWLEDGE).

This very crude overview of inferentialist theories shows how far-reaching the implications of a theory of concept individuation/possession can be. The possibility of employing inferentialism to provide an account of the a priori and ground conceptual analysis is obviously one of the most appealing features of the view. But the view also has other far-reaching consequences, ones that can be used against it. A famous case recently proposed by Williamson (2007) will illustrate the point⁹. Peter is a leading logician with “deviant” views. He rejects the sentence:

- (1) Every vixen is a female fox

for the following reasons. First, he takes universally quantified statements to be existentially committing, so he thinks the claim is false unless there is at least one vixen. (Suppose he has developed a sophisticated semantic theory in support of his view). In addition, Peter believes there are no foxes, and thus no vixens, our credence to the contrary being induced by mass hallucinations provoked by the government.

Now, when *we* assert (1) we express the thought EVERY VIXEN IS A FEMALE FOX. Does Peter *reject* that very same thought when he asserts the negation of (1)? If so, Peter has the concept VIXEN (since he rejects a thought constituted by it) *without* being disposed to accept X IS A VIXEN → X IS A FEMALE FOX. This is incompatible with an inferentialist theory of concept individuation/possession, on which (presumably) one must be disposed to make that inference in order to have VIXEN (cf. our previous example involving BACHELOR).

The inferentialist must therefore *deny* that Peter expresses the same thought as us with his utterance: he expresses (and rejects) a *different* thought constituted by a *different* concept VIXEN* (something like: EVERY VIXEN* IS A FEMALE FOX) But:

[...] the linguistic understanding of (1) we share with Peter [...] already suffices for [him] and us to articulate our disagreements in rational discourse; we are not merely talking past one another. In its small way, (1) determines a piece of the common intellectual heritage of mankind, something we share with Peter [...] in our very capacity to disagree over it. To insist that the thought we associate with (1) nevertheless differs from [the thought Peter associates] with (1) is to undermine Frege’s requirement of the publicity of senses, and in particular thoughts (Williamson 2007, p. 114).

⁹ V. Williamson 2007, pp. 73-121. Williamson’s original case is slightly different from the one I use in the main text. (In fact, Williamson’s case is more radical. Williamson also employs my version of the case on pp. 116-117).

On an inferentialist theory, deviant subjects who reject those fundamental inferences which are allegedly required for concept possession will turn out to have different concepts from us (VIXEN* instead of VIXEN). Consequently, they will also express different thoughts when using the relevant linguistic expressions. According to anti-inferentialists like Williamson, this makes it hard to account for the fact that we can still *agree*, *disagree* and *communicate* with them when using such expressions, since this seems to require sameness of expressed thought¹⁰. In this way, the debate on concept individuation and possession connects with a number of further issues, such as the conditions for communication and agreement/disagreement in philosophy of language, or the nature of theory change in philosophy of science¹¹.

Determining whether there are any inferential conditions on concept individuation/possession, and whether this would preclude the “shareability” or “publicity” of concepts, will be one of the main goals of our discussion. As our introductory remarks show, any theory of concept individuation and possession will face the following questions:

- Is it the case that every concept is individuated by a set of *inferential connections* which are required for its possession?
- If so, *which* of the connections in which a concept actually stands are part of its individuation conditions and are required in order to have the concept?

Three main positions on the issue have been occupied in the literature. All of them will be extensively discussed in the following chapters, but a general overview will be useful. As we have seen, “inferentialist” or “Inferential Role Semantics” (IRS) views¹² answer the first question affirmatively. Concepts *are* inferentially individuated: for every concept C, there is a set of inferential dispositions which are required in order for someone to have C. For instance, one doesn’t have BACHELOR unless he is disposed to infer from X IS A BACHELOR to X IS AN UNMARRIED MAN, and vice versa.

This raises our second question for IRS. Any given concept will stand in many different inferential connections. For instance, I might think all bachelors are sad, in which case I will be disposed to infer from X IS A BACHELOR to X IS SAD. Is my concept BACHELOR also individuated by the inference BACHELOR → SAD? More generally: is a concept individuated by *all* the inferential connections in which it stands, or by only some of them? Depending on how they answer this question, we can distinguish two main varieties of IRS views. According to *holistic* versions of IRS, a concept is

¹⁰ Again, the literature on the topic is extremely vast: v. e.g. Sainsbury and Tye 2012 (p. 21) and Laurence and Margolis 2003 (p. 262) for a similar line. (We will go back to this in ch. 6).

¹¹ Cf. e.g. the remarks of Fodor and Lepore (1992, pp. 11-13) on Kuhn (1962).

¹² This is the most common label for the view (v. e.g. Fodor 1998) and the one I will usually employ in what follows.

individuated by *all* of its inferential connections¹³. It follows that, in order for you to have my concept BACHELOR, you must not only be disposed to infer X IS A BACHELOR → X IS AN UNMARRIED MAN (and vice versa), but also X IS A BACHELOR → X IS SAD. (Indeed, you will need to share all of my bachelor-related beliefs, so it will be very hard for you to have my concept BACHELOR!).

According to *non-holistic* (or *localist*) versions of IRS, only *some* of a concept's inferential connections will be included in its individuation conditions¹⁴. For instance, the inferences X IS A BACHELOR → X IS AN UNMARRIED MAN and X IS AN UNMARRIED MAN → X IS A BACHELOR do individuate the concept, while X IS A BACHELOR → X IS SAD does not; consequently, only the former inferential dispositions are required for you to have the same concept BACHELOR as me.

Another important family of views will just answer our first question in the negative: according to *atomistic* theories, concepts are *not* individuated inferentially, and *no* specific inferential dispositions are required for their possession¹⁵. For instance, one could have the concept BACHELOR even if he was not disposed to infer X IS A BACHELOR → X IS AN UNMARRIED MAN. Similarly, Williamson's "deviant logician" can have the concept VIXEN even though he does not accept X IS A VIXEN → X IS A FEMALE FOX because of his bizarre views¹⁶.

The following chapters will be, to a large extent, a discussion of these three positions regarding the role of inferences in concept individuation/possession. Indeed, the main goal of my work can be summarized as follows:

¹³ V. ch. 2 (sect. 2.3) for more details about holistic IRS.

¹⁴ Ch. 6 will focus almost entirely on localist views, which will also be discussed in chs. 4-5.

¹⁵ In itself, atomism is a purely negative position: it simply tells us that concepts are *not* individuated by their inferential connections. Defenders of the view have then supplemented their main claim with a positive account of concept individuation (v. for instance Fodor's "Informational Atomism", which will be discussed in chs. 4-5).

¹⁶ Many atomists restrict their claims to *lexical* concepts, i.e. concepts expressed by single morphemes in natural language. For instance, Fodor holds that FEMALE FOX is a complex concept constituted by FEMALE and FOX (v. e.g. Fodor 2008, p. 59). Therefore, he grants that someone has FEMALE FOX only if he is disposed to infer from X IS A FEMALE FOX to X IS A FOX. (This has to do with his views on compositionality: v. Fodor and Lepore 2002). At the same time, he thinks someone could have VIXEN without being disposed to infer from X IS A VIXEN to X IS A FOX, since VIXEN is a lexical, non-complex concept.

This is a tricky issue. The notion of "lexical concept" is not easy to define ("mailman" is a single word, but the concept it expresses is arguably complex (Weiskopf 2009a, fn. 2)). Moreover, a concept might be expressed by a single morpheme in a language and by a complex phrase in another. This makes the restriction to lexical concepts hard to spell out. Finally, some atomists (Williamson) would *deny* that, in order to have FEMALE FOX, one must be disposed to infer from X IS A FEMALE FOX to X IS A FOX (v. Williamson 2007, pp. 73-116). So the restriction to lexical concepts is controversial. For these reasons, I will simply take atomism to apply to concepts in general rather than lexical concepts: nothing will turn on this, and the reader should feel free to restrict the claims of atomists like Fodor to lexical concepts if needed.

- Determining whether there are any inferential constraints on concept individuation and possession.
- Developing the foundations for a substantive theory of concept individuation/possession.

Both my approach to the problem and my solution to it present several distinctive features. It will then be useful for the reader to have a general idea of the view I am going to develop, and of how I will argue for it throughout the following chapters.

Within contemporary debates on concepts, it has become a somewhat standard methodology to list a set of independently plausible “constraints” or “*desiderata*” that we want a good theory of concepts to satisfy; this is then used to show that a certain view is superior to its competitors in its ability to satisfy most or all of these constraints¹⁷. **Chapters 1-3** will be aimed at discussing one of these constraints, the so-called “publicity” principle. According to this principle, concepts are public entities which are routinely shared by ordinary thinkers. In particular, I will focus on a version of the principle (defended by Fodor and others) on which concepts must be shared in order for interpersonally applicable *psychological generalizations* to be possible. The principle has played a crucial role in recent debates on concepts; in particular, it has been frequently used by anti-inferentialists like Fodor to argue against holistic versions of IRS. This makes it crucial for anyone interested in issues of concept individuation/possession to establish whether publicity is true, and whether it is indeed incompatible with holism.

I will suggest a new approach to the problem of publicity¹⁸. My driving hypothesis is that we should look at intentional generalizations (e.g. “If someone wants water, then he will look for water other things being equal”) as *sentences in a language*. Such sentences embed *propositional attitude ascriptions* (“If someone wants water...”): the conditions for them to apply to a group of subjects will therefore depend on the content of those ascriptions. We cannot establish whether a generalization requires the subjects it covers to share certain specific concepts (e.g. the concept WATER) until we get clearer on its semantics. For this reason, our discussion of concept individuation and possession will have an unusual starting point. **Chapter 1** will be an in-depth examination of so-called “Millian” views about the semantics of reports. First, I will present a set of arguments against the “non-pragmatic” Millian approach recently developed by David Braun. Then, I will review some of the standard objections against classic “pragmatic Millian” accounts (Salmon, Soames). Finally, I will offer some reasons to think that the truth conditions of attitude reporting sentences vary with context. My conclusion will be that a satisfactory theory of ascriptions must be *contextualist* and *non-Millian*, two *desiderata* that will play a key role in our later discussion.

¹⁷ V. e.g. Fodor (1998), Prinz (2002).

¹⁸ My approach was influenced by Schneider (2005, 2011): v. ch. 3 (sect. 4.2) for more details.

In **chapter 2**, I will present the standard arguments for publicity and the anti-holistic objections that have been based on them. I will reject two popular holistic responses to such objections: the appeal to fundamental intentional generalizations and the appeal to concept similarity. Then, I will try to state the standard arguments for publicity in a more precise, rigorous form. This will make it easier to show that such arguments are affected by a previously unnoticed problem, and that we can only make them valid by supplementing them with some highly controversial additional premises.

Having highlighted the problems with standard arguments, I will present a new argument for publicity and a new version of the principle in **chapter 3**. There, I will discuss those *contextualist* theories of attitude reports (Crimmins and Perry, Richard, Forbes) that have been specifically designed to meet the two *desiderata* laid out in ch. 1. Such theories have immediate consequences for the semantics of intentional generalizations. Once we accept contextualism, such generalizations will encode information about the *concepts* under which intentional agents relate to the relevant propositions, and not only about the propositions themselves (as on Millian views). In turn, this establishes a contextualist version of publicity: the new principle requires the subjects to which a generalization applies (in a given context) to share the specific concepts that were “selected” by the speaker (in that context) for the purpose of his generalization.

I will then argue that the principle is still incompatible with existing holistic theories (Block, Schneider). However, I will also sketch a new, more moderate holistic position which *would* satisfy publicity. On that view, *some* concepts are holistically individuated and not shared, while others are individuated more coarsely and possessed by multiple thinkers. This revised holistic view will then be the core of the broader “pluralist” picture of concept individuation/possession which I will defend in the following chapters.

In **chapters 4-5**, I turn to examine the relationship between publicity and a second widely endorsed constraint on a theory of concepts, the so-called “Fregean Constraint” (FC). (FC) claims that subjects who find themselves in a “Frege case” and mistakenly ascribe contradictory properties to the same object must have *distinct* concepts for that object. I will argue that there is a tension between (FC) and publicity, and that none of the theories of concepts currently on the market seems to satisfy both constraints: all those theories that individuate concepts finely enough to satisfy (FC) also appear unable to account for their shareability. While some issues in the vicinity have been noted, no one in the literature has (to my knowledge) stated the tension between our two constraints in these terms. This is unfortunate, for that tension can teach us a lot about how to develop a theory of concept individuation and possession.

I will discuss some possible ways to satisfy both of our constraints at the same time. I will not so much try to argue for a specific solution to our problem, but rather to show that, on *all* possible solutions, a certain theory of concept individuation/possession becomes overwhelmingly plausible. First, I will argue that the moderate holistic view sketched at the end of ch. 3 should be part of our theory of concepts if we want to

satisfy (FC). Then, I will integrate that view in a more general picture of concept individuation/possession. This is the picture of concepts that appears most plausible in light of our two constraints, since it's the one that does the best job at solving the tension between them¹⁹.

Here is a brief overview. On my picture, there isn't a *single* answer to questions of concept individuation and possession: different concepts are individuated in different ways and have different possession conditions. My approach is thus very different from that of atomist, localist and holistic theories. I take each of these views to provide a correct account of the individuation/possession conditions of *some*, but not *all* concepts. My picture integrates all three views, without fully endorsing any of them. I take *some* concepts to be holistically individuated: holism gives us the fine-grained individuation conditions we need to satisfy (FC) in all the relevant Frege cases. At the same time, some *other* concepts must be individuated non-holistically: this is required if we want to account for concept sharing and satisfy the publicity principle established in chapter 3. For this reason, I take some concepts to be individuated by a small set of their inferential connections (localism), and I am open to the possibility of individuating some concepts without appealing to their inferential connections *at all* (atomism).

My *pluralist* theory of concepts has some features in common with pluralist views such as Weiskopf (2009 a,b), but it also bears some crucial differences with them²⁰. First, my view is significantly more radical in its pluralism. Weiskopf still takes all concepts to be locally individuated, while I think we should allow for holistically individuated concepts and, quite possibly, for concepts that are not individuated inferentially at all. Second, Weiskopf's only argument for pluralism is based on its ability to account for recent findings in cognitive psychology. My argument is very different: I think fit with empirical data is an important bonus for pluralist pictures, but their main strength lies elsewhere. Concepts must be individuated in multiple ways because this is the only way for them to play two of their roles: on the one hand, accounting for the interpersonal applicability of intentional generalizations (= publicity); on the other, explaining how subjects involved in Frege cases can be rational while ascribing contradictory properties to the same object (= FC). Finally, a further distinctive feature of my view is that it brings out for the first time the role of *context* in a theory of concepts. If concepts are individuated in a variety of different ways, it will be speakers' intentions to determine *which* specific concepts are involved in our intentional generalizations and explanations: in accordance with the theory of ascriptions defended in chapter 3, speakers will "select" more coarsely individuated concepts in certain contexts, and more finely individuated concepts in others (v. ch. 5, sect. 4.3 for more details).

¹⁹ I will not discuss certain other constraints that have also played an important role in the debate on concepts, such as compositionality (Fodor 1998, 2008; Fodor and Lepore 1992, 1996, 2002). I believe my theory would also meet these constraints, but here I prefer to focus on publicity and (FC).

²⁰ I compare the two views in ch. 6 (sect. 2.3).

In **chapter 6**, I conduct an in-depth discussion of various forms of localism about concepts. This is not only interesting in itself; it also serves the purpose of spelling out my view further by contrasting it with a family of rival theories. I present a set of dilemmas for localism and consider various localist responses to each of them. The main focus is on the *pluralist-localist* position defended by Weiskopf (v. *supra*) and the *descriptivist-localist* position defended by Jackson (1998). On Jackson's view, the inferences which individuate a concept and are required for its possession *fix the reference* of that concept. First, I will argue that most localists are in fact committed to Jackson's descriptivism. Then, I will develop an objection by Schroeter (2004) to show how the view makes wrong predictions about the reference of natural kind concepts. If reference is fixed by our inferential dispositions, subjects whose dispositions are sufficiently "deviant" will turn out to refer to *different natural kinds* from us; in turn, this will have deeply counterintuitive consequences in all those cases where deviant subjects appear to be theorizing about the very same natural kinds as us.

Chapter 1

Two Varieties Of Millianism

1. Introduction

As anticipated, our discussion has a slightly unexpected starting point. This chapter will critically assess a family of views about the semantics of *propositional attitude ascriptions*, which goes under the name of “Millianism”²¹. While interesting in itself, my discussion of Millianism will also have crucial implications for the problem of concept individuation and possession, for reasons that will become apparent in chapter 3.

Millianism is one of the main theories of attitude reports currently on the market. Millian views have been developed by (among others) Nathan Salmon (1986, 1989, 2006), Scott Soames (1987, 1995, 2002) and David Braun (1998, 2000, 2001a,b). While sharing a core set of assumptions, these theories also present some crucial differences. In particular, we can distinguish between “pragmatic” (Salmon, Soames) and “non-pragmatic” (Braun) versions of Millianism. While pragmatic accounts have been far more popular, they also seem to be affected by several problems, which have pushed authors like Braun to develop their non-pragmatic alternative.

Having described the basic principles of Millianism (sect. 2), I will move on to examine Braun’s non-pragmatic view (sects. 3-4). I will try to determine whether the view can deal with two classic objections that all Millians have to face. The first objection is that, if Millianism is true, then speakers who are clearly rational will often have contradictory beliefs because of their intuitions about belief reports; the Millian must then explain how these speakers can be rational even though they believe a contradiction. Against Braun, I will argue that Non-Pragmatic Millianism is unable to explain the rationality of ordinary speakers who are aware of certain identity facts. A similar objection has been raised by Stephen Schiffer (1987, 1990, 2006); I will first argue that Braun’s response to the objection doesn’t work, and then that there is no way for Non-Pragmatic Millianism to offer an alternative solution to the problem while holding on to its basic commitments.

A second classic objection against Millianism is that the view cannot account for the role that certain intentional generalizations play in the explanation and prediction of behavior. Braun (2000, 2001a,b) has proposed a non-pragmatic solution to the problem; in reply, I will show that his strategy still fails to account for a large number of powerful psychological generalizations, and that, again, the only way in which the Millian can hope to account for such generalizations is to appeal to what they pragmatically convey.

²¹ The view sometimes goes by other names: “Russellianism” (Braun 1998, Richard 1990), “Naive Russellianism”, “The ‘Fido’-Fido theory” (Schiffer 1987) or “The Direct Reference Theory”. I find “Millianism” to be the least misleading one and I will stick to it in what follows. Also, note that Millianism is sometimes identified with a theory about the content of proper names: the view I call “Millianism” includes a Millian theory of names and supplements it with further assumptions regarding the semantics of proper names within attitude contexts (v. sect. 2.1 *infra*).

In light of these two objections, my conclusion will be that Non-Pragmatic Millianism should be rejected. The only way for the Millian to explain speakers' rationality and account for intentional generalizations is to adopt a pragmatic view: *if* Millianism is true, *then* Pragmatic Millianism is. This, of course, has a further important consequence: if Pragmatic Millianism also fails, then Millianism must be rejected altogether. And, indeed, pragmatic accounts do have to face some extremely powerful objections, which I will review in section 5: from this, I will conclude that a successful theory of ascriptions must be *non-Millian*.

I will then end the chapter by discussing a second feature that a theory of ascriptions should take into account: the truth conditions of attitude-reporting sentences seem to *vary with context*. In chapter 3, we will see how these two *desiderata* have consequences that extend far beyond the problem of ascriptions, touching directly on issues of concept individuation and possession.

2. Millianism: An Overview

2.1. Millian Theories And The Substitution Objection

What is a Millian theory of propositional attitude ascriptions? Following Braun (1998, pp. 557-558), we can summarize the fundamental principles of Millianism as follows:

a) Propositions:

- Propositions are the semantic contents of sentences and are expressed by them.
- A proposition is a structured entity whose constituents are the contents of the parts of the sentence expressing the proposition in question.
- The content of a predicate is a property or relation; the content of a proper name or indexical is the individual to which the name or indexical refers.

It follows from this set of Millian principles that the sentences:

- 1) Superman can fly.
- 2) Clark Kent can fly²².

express the *same* proposition, whose constituents are the individual Superman/Clark Kent and the property of *being able to fly*. This proposition can be represented thus:

<Superman, *being able to fly*>

Alternatively, it can be represented thus:

<Clark Kent, *being able to fly*>

(Following standard usage, I will refer to propositions having individuals and properties as their constituents as “Russellian propositions”).

b) Belief reports:

- The content of the predicate “believes” is a binary relation which is said to obtain between a subject and a proposition.
- The content of a “that”-clause “that S” is the proposition normally expressed by S.

²² V. Appendix A for a list of numbered sentences for ch. 1.

It follows from our two sets of Millian principles that the sentences:

- 3) Lois Lane believes that Superman can fly.
- 4) Lois Lane believes that Clark Kent can fly.

express the *same proposition* (whose constituents are Lois Lane, the proposition that Superman/Clark Kent can fly, and the belief relation) and must therefore have *same truth value*. This Russellian proposition can be represented thus:

<Lois Lane, <Superman, *being able to fly*, *believing*>

Alternatively, it can be represented thus:

<Lois Lane, <Clark Kent, *being able to fly*, *believing*>

Millianism faces an obvious worry; following Braun (1998), I will call it “the Substitution objection”. Competent speakers who are familiar with the Superman story and reflect carefully on the belief reports (3)-(4) will mostly take them to have different truth values; in particular, they will take (3) to be true and (4) to be false. If asked to justify their judgments, they will point out that according to the story Lois would assent to “Superman can fly” but dissent from “Clark Kent can fly”. Therefore, they will conclude, Lois believes that Superman can fly but does not believe that Clark Kent can fly²³.

Let’s call “anti-substitution intuitions” the judgments that (3)-(4) have different truth values. Such intuitions pose a problem for the Millian. It seems that ordinary speakers are not being irrational in thinking (falsely, according to Millianism) that (3) is true and (4) false. But, presumably, if someone believes that a certain sentence is true then he believes whatever proposition is expressed by that sentence. Conversely, if someone believes that a certain sentence is false then he believes the negation of whatever proposition is expressed by that sentence. So a speaker who thinks that (3) is true and (4) false thereby believes the proposition expressed by (3) and the negation of the proposition expressed by (4). But if Millianism is true (3) and (4) express the same proposition. Therefore, if Millianism is true most ordinary speakers *rationally* believe a proposition and its negation, i.e. they rationally believe a contradiction. But how can someone believe a contradiction and yet be rational? The Millian must *explain* how competent and reflective speakers who understand (3)-(4) can rationally take them to have different truth values; he must specify what makes such speakers rational despite their contradictory beliefs.

²³ In his (1998), Braun uses a Hesperus/Phosphorus case; I prefer to employ a Superman/Clark Kent case to avoid various complications.

2.2. A Millian Metaphysics Of Belief

Speakers don't have anti-substitution intuitions about attitude reports only; in fact, the Millian seems to have exactly the same problem with simple sentences. Suppose Lois is a competent speaker who reflects carefully on the simple sentences (1)-(2). Like most speakers in the Superman story, Lois thinks (1) is true and (2) false, and it seems she is not irrational in making this (false) judgment. Assuming Millianism, however, it follows by the same argument considered in the last section that Lois believes a proposition and its negation, i.e. a contradiction. How, then, can she be rational in her judgments about (1)-(2)?

Millians generally deal with anti-substitution intuitions about simple sentences by appealing to a certain “metaphysics of belief”, originally developed by Salmon (1986). This metaphysics relies on the notion of a “way” or “guise” under which a certain proposition is believed or disbelieved²⁴. On a Millian metaphysics, Lois believes the proposition expressed by (1)-(2) *while taking it in one way* and believes the negation of that very same proposition *while taking it in a suitably different way*. In particular, she believes the proposition expressed by (1)-(2) when it's “presented to her” under the “guise” of (1), but believes its negation when the same proposition is presented to her under the guise of (2). Since she thinks that the referent of “Superman” is different from the referent of “Clark Kent”, she is not aware that (1)-(2) are in fact guises of one and the same proposition. Therefore, given her evidence about Superman/Clark, she judges the two sentences to have different truth values, thus ending up with a contradictory belief. Because she doesn't know that (1)-(2) are guises of the same proposition, however, the ways in which she believes and disbelieves that proposition are “suitably different” and she is not irrational.

Summarizing: on the Millian metaphysics developed by Salmon, a subject S believes a proposition P just in case S stands in an appropriate “acceptance” relation (which Salmon calls “BEL”) to a way w expressing P. The truth conditions of a belief report will then be as follows (where BEL is a three-place relation holding between a subject, a proposition and a way of believing that proposition, and “ w ” is a variable ranging over ways of believing)²⁵:

$$S \text{ believes that } P \text{ iff } (\exists w) (\text{BEL}(S, P, w))^{26}$$

²⁴ A subject “disbelieves” a proposition just in case he believes its negation.

²⁵ Adapted from Salmon (1986, p. 111).

²⁶ Notice that no reference is being made to a *specific* way under which the proposition in question is believed by the subject. This marks the difference between “true” Millian theories and views such as Crimmins and Perry (1989) or Richard (1990), which also take “that”-clauses to express Russellian propositions. (We will discuss these views in ch. 3 *infra*).

For instance, Lois believes the proposition that Superman/Clark can fly in a way corresponding to the natural language sentence (1), while believing its negation in a way corresponding to the negation of the natural language sentence (2)²⁷.

2.3. Pragmatic Millianism

Millians like Salmon, Soames and Braun *all* endorse the metaphysics of belief just described and generally *agree* that it provides a satisfactory account of intuitions about simple sentences like (1)-(2). However, Non-Pragmatic Millians (NPM) *also* think that metaphysics is enough to explain the rationality of anti-substitution intuitions about reports like (3)-(4), while Pragmatic Millians (PM) invoke a pragmatic mechanism to account for the latter set of judgments²⁸. According to the Pragmatic Millian, utterances of (3)-(4), while semantically expressing the *same proposition*, pragmatically convey *different propositions* which *do* have different truth values. These propositions can be construed in different ways. According to Salmon, utterances of (3)-(4) might systematically convey propositions concerning the guises under which Lois believes a certain proposition (Salmon 1986, pp. 115-118; Salmon 1989, pp. 248-250):

- 5) Lois Lane believes the proposition that Superman/Clark Kent can fly under the guise of the sentence “Superman can fly”.
- 6) Lois Lane believes the proposition that Superman/Clark Kent can fly under the guise of the sentence “Clark Kent can fly”.

Soames (2002) proposes a different account, on which speakers systematically convey “descriptively enriched” propositions through their utterances of (3)-(4). For instance, speakers might use (3)-(4) to convey propositions like:

- 7) Lois Lane believes that Superman, the mighty superhero, can fly.
- 8) Lois Lane believes that Clark Kent, the milquetoast reporter, can fly.

On both accounts, utterances of (3) pragmatically convey a *true* proposition (the one expressed by (5) or (7)), while utterances of (4) pragmatically convey a *false* proposition (the one expressed by (6) or (8)). Pragmatic Millians will then explain speakers’ anti-substitution intuitions in (roughly) the following way:

- a) Speakers mistakenly take the propositions conveyed by utterances of (3)-(4) to constitute the semantic content of (3)-(4).
- b) Speakers believe the (true) proposition conveyed by utterances of (3) and disbelieve the (false) proposition conveyed by utterances of (4).

²⁷ Later on, we will consider some alternative construals of ways of believing.

²⁸ I borrow this taxonomy from Braun (1998), who uses a different terminology.

- c) Therefore, they conclude that (3) is true and (4) false, i.e. that (3)-(4) have different truth values.

2.4. Non-Pragmatic Millianism

Pragmatic accounts of anti-substitution intuitions have been extremely popular among Millians, but they also have to face several objections (v. sect. 5 *infra* for an overview). This has pushed Millians like Braun to *reject* pragmatic responses to the Substitution objection and argue for a form of “Non-Pragmatic Millianism” (NPM). As we have seen, *all* Millians agree that the rationality of intuitions about *simple sentences* can be explained by simply appealing to a certain metaphysics of belief and without invoking any pragmatic mechanisms; Braun thinks that the Millian should extend this account to attitude reports. More specifically: according to NPM, we can offer a *psychological* explanation of why speakers are rational in their intuitions about (3)-(4); unlike non-Millian views, this explanation will not assume that (3)-(4) have different semantic contents, and unlike Pragmatic Millian views it will *also* not assume that utterances of (3)-(4) pragmatically convey different propositions (Braun 1998, p. 579).

Braun’s response to the Substitution objection is supposed to rely entirely on the Millian metaphysics of belief developed by Salmon (1986), with one important difference. Braun identifies the “ways” or “guises” through which a subject believes a proposition with sentences in the Language of Thought (LoT)²⁹: a subject S believes a proposition P just in case he has a mental representation (a “mental sentence”) expressing P and playing the functional role appropriate for belief. In this case, we say that S has the mental sentence in question in his “belief-box” and that he thereby believes that P.

We have already seen (sect. 2.2) how a Millian metaphysics of belief can be used to account for intuitions about simple sentences. Lois assents to (1) and dissents from (2), thus believing a proposition and its negation. On Braun’s account, she believes that proposition by having the mental sentence SUPERMAN CAN FLY – which corresponds to (1) – in her belief-box; and she disbelieves that proposition by having the mental sentence CLARK KENT CANNOT FLY – which corresponds to the negation of (2) (call it (2n)) – in her belief-box³⁰. Since she has (1) and (2n) in her belief-box, she is disposed to assent to (1) and dissent from (2). She thus believes and disbelieves the very same proposition, but she is not irrational because she does so in *suitably different ways*, i.e. by having *suitably different LoT sentences* in her belief-box.

²⁹ More precisely: although Braun does think (for Paderewski-related reasons) that we should identify ways with mental, non-linguistic representations, he is open to non-LoT construals of the mental representations in question.

³⁰ Following Braun, I assume for the sake of simplicity that natural language sentences can work as mental sentences in LoT.

According to Braun, this psychological account of anti-substitution intuitions about simple sentences can be straightforwardly extended to intuitions about reports. Suppose Peter, a colleague of Lois, has seen her sincerely assent to (1) and dissent from (2); as a result, he ends up accepting (3) and denying (4), i.e. accepting the negation of (4) (call it (4n)). If Millianism is true, Peter believes a contradiction; still, we can account for his rationality just like we did for Lois. Peter has the equivalent of (3) and (4n) in his belief-box: he is therefore disposed to assent to (3) and dissent from (4). Moreover, the ways (i.e. the LoT sentences) through which he believes and disbelieves the proposition expressed by (3)-(4) are *suitably different* and he is therefore not irrational.

NPM thus offers a *unified* psychological explanation for the rationality of anti-substitution intuitions about simple sentences and reports. I will summarize this explanation through the following schema, which applies (*mutatis mutandis*) to both Lois and Peter³¹:

(Sch): *If* a competent, reflective speaker S judges that a sentence (α) is true and a sentence (β) is false, and:

- (i) (α)-(β) express the same proposition P;
- (ii) S believes the proposition P expressed by (α)-(β) by having the LoT equivalent of (α) in his belief-box;
- (iii) S disbelieves the proposition P expressed by (α)-(β) by having the LoT equivalent of not-(β) in his belief-box;
- (iv) The LoT equivalents of (α) and not-(β) are *suitably different ways* in which S believes and disbelieves P;

Then, other things being equal³², S is rational³³.

What is it for two LoT sentences to be suitably different ways of taking a certain proposition? There are at least two possible construals of the notion:

Option A: This is the option Braun seems to favor:

[Lois] would not be able to deduce any contradiction from [the propositions expressed by (1) and (2n)], given the ways in which she believes them, for no contradictory sentence can be validly derived (in the syntactic sense) from sentences (1) and (2n) alone. Thus it may be no more irrational for her to have (1) and (2n) in her belief-box than for her to have ‘Gingrich is a Republican’ and

³¹ The schema is inspired to Schiffer’s “Frege’s Constraint” (Schiffer 2006, p. 362).

³² This rules out clearly *irrational* subjects who *do* satisfy (i)-(iv), but who (say) also believe that (α) and not-(α) are both true, or who believe that (α) is true and (β) false for no reason, etc... Thanks to the “*ceteris paribus*” clause, (Sch) will not count these subjects as rational. Notice that neither Lois nor Peter is obviously irrational in any of these ways, so (Sch) applies to both of them.

³³ More precisely: S’s judgment (that (α) is true and (β) false) is rational. I will leave this qualification implicit in what follows.

‘Clinton is not a Republican’ in her belief-box (Braun 1998, pp. 575-576. In this quote and the following ones I made some changes from Braun’s original text, in order to conform to my own choice of examples).

On this construal, two LoT sentences are suitably different for S just in case S cannot “syntactically” derive a contradictory sentence from them.

Option B: Alternatively, the NPM theorist might hold that two LoT sentences are suitably different for S just in case S takes the corresponding natural language sentences to express different propositions. Like the first option, this would also enable us to explain Lois’ rationality through (Sch), for Lois takes “Superman” and “Clark Kent” to refer to different individuals and therefore also takes (1) and (2) to express different propositions.

3. NPM And Enlightenment: When Psychology Is Not Enough

3.1. The Asymmetry Objection

In this section, I will argue that Braun's psychological, non-pragmatic account of anti-substitution intuitions fails. More specifically, I will argue that his appeal to ways of believing, far from making PM's pragmatic explanation useless, will in fact *require* invoking pragmatic mechanisms in order to account for speakers' rationality. This shows that NPM's attempt to explain anti-substitution intuitions in a purely psychological fashion fails, and that the Millian will only be able to explain those intuitions if some PM account can be made to work.

Suppose Peter is “enlightened” about Superman’s secret identity, i.e. suppose he accepts the following identity sentence:

9) Superman is Clark Kent.

If enlightened Peter is like most ordinary speakers, he will still take (3) to be true and (4) to be false, and it seems he will still be rational in doing so. How can NPM explain the rationality of his judgments?

Braun (1998) himself considers this possible variant of the Substitution objection³⁴; his response is that NPM can account for Peter’s rationality in the same way as before, i.e. by appealing to his having suitably different ways of believing the relevant proposition:

The explanation is essentially the same as before. He believes the proposition expressed by (3) in one way, a way corresponding to (3); he fails to believe it in another way, a way corresponding to (4). In fact, he believes the negation of the proposition in a way that corresponds to (4n). [...] Therefore, he believes that (3) and (9) are true, and (4) is false (Braun 1998, p. 583, emphasis mine).

We can see why Braun’s response fails by considering the following problem, which I will name “the Asymmetry objection”. (Braun does consider this potential objection³⁵; my way of presenting the problem is inspired by his, but there are also many substantial differences between the two. I find Braun’s description of the issue partially unsatisfactory, for reasons I will not discuss here. Braun also offers a reply to the objection, which I will discuss in sect. 3.3 (v. especially fn. 44)).

Suppose Randy, a colleague of Peter and Lois, is also “enlightened”, i.e. he also accepts (9). Moreover, assume that:

³⁴ This variant was originally brought up by Schiffer as a problem for Millian theories in general: cf. Schiffer (1987, pp. 463-466).

³⁵ Cf. Braun (1998, pp. 583-590). A similar objection is raised by Schiffer (1987, p. 465).

- Randy is as reflective and as competent a speaker as Peter;
- Randy is exactly like Peter in philosophical background, training on logic and semantics and so on. In short, they are both equally average speakers under all relevant respects.

Finally, suppose Randy carefully evaluates (1)-(2) and claims that (1) is true but (2) false. That is, suppose he resists substitution with respect to the simple sentences (1)-(2) while also accepting the identity claim (9). Randy seems *irrational*: (2) is a “[...] *simple* logical consequence of (1) and (9)” (Braun 1998, p. 584) and it can be easily derived from them by substituting coreferential terms, so it does not seem rational to accept (1)-(9) but reject (2) after carefully reflecting on the relevant identity facts.

Now, if Millianism is true (4) is *also* a simple logical consequence of (3)-(9) and it can *also* be derived from them by substitution of coreferential terms. So, if Millianism is true, someone who accepts (3)-(9) while taking (4) to be false should also count as irrational. But enlightened Peter doesn’t seem irrational, even though (like most enlightened speakers who are familiar with the Superman story) he accepts (3)-(9) and rejects (4).

Enlightened Randy is irrational but enlightened Peter is not³⁶; if the Millian grants this, he must somehow explain their difference in rationality. This is a problem for Braun’s psychological account, for I will argue that Peter and Randy are *exactly alike* in their ways of taking the respective propositions: if Peter has suitably different ways of believing and disbelieving the proposition expressed by (3)-(4), so does Randy with respect to the proposition expressed by (1)-(2). In other words, I will argue that the following conditional is true:

(C): *If* the LoT equivalents of (3) and (4n) are *suitably different ways* in which Peter believes and disbelieves that Lois believes that Superman/Clark Kent can fly, *then* the LoT equivalents of (1) and (2n) are *suitably different ways* in which Randy believes and disbelieves that Superman/Clark Kent can fly.

(C) seems true on both possible construals of the notion of “suitably different” ways of believing:

Option A: According to this option, which Braun favors, two LoT sentences are suitably different for S just in case S cannot “syntactically” derive a contradictory sentence from them. Now, according to NPM Peter has the LoT sentences (3), (4n) and (9) in his belief-box. So NPM will explain why Peter is rational by claiming that he cannot syntactically derive a contradiction from the inconsistent sentences (3)-(4n), even though he also accepts (9). But if this is true of Peter, then Randy too cannot

³⁶ From now on I will simply refer to them as “Peter” and “Randy”, taking for granted that they are both enlightened.

syntactically derive a contradiction from (1)-(2n); the pairs of LoT sentences (1)-(2) and (3)-(4) are exactly analogous and there is no reason to suppose that a contradiction could be syntactically derived from (1)-(2n)-(9) but not from (3)-(4n)-(9). So under option A Randy has suitably different ways if Peter does; therefore, we cannot explain their difference in rationality by claiming that only Peter has suitably different ways of taking the relevant proposition.

Option B: According to this option, two LoT sentences are suitably different for S just in case S takes the corresponding natural language sentences to express different propositions. For instance, Lois, who is unenlightened, takes (1) and (2) to express different propositions because she takes “Superman” and “Clark Kent” to refer to different individuals. Therefore, she is rational in accepting (1) and (2n). But Peter and Randy are both enlightened; therefore, NPM cannot claim that they take their respective pairs of sentences to express different propositions because of their taking “Superman” and “Clark Kent” to refer to different individuals. And if Peter takes (3)-(4) to express different propositions for some *other* reason, then Randy will *also* take (1)-(2) to express different propositions³⁷, since they are *ex hypothesi* exactly alike under all relevant respects (they are both average competent speakers, etc...). So under option B, too, Randy has suitably different ways if Peter does, which again blocks NPM’s psychological explanation.

On both possible interpretations of the notion of “suitably different ways”, (C) is true: if Peter has suitably different ways, so does Randy; conversely, if Randy lacks suitably different ways, so does Peter. Therefore, NPM cannot explain why Peter is rational and Randy irrational by claiming that Peter, but not Randy, has suitably different ways of believing/disbelieving the relevant proposition.

Less informally: *ex hypothesi*, both Peter and Randy satisfy conditions (i)-(iii) in (Sch); given (C), if Peter satisfies (iv) then Randy also does; so, if Peter satisfies (i)-(iv), then Randy does too. Therefore, we cannot explain why Peter is rational by assuming (Sch) and claiming that he satisfies (i)-(iv), for then it will follow that Randy is also rational. But NPM claims that (Sch) explains the rationality of anti-substitution intuitions about reports; therefore, NPM is false.

3.2. First Reply: Biting The Bullet

A first possible reply would be to deny the asymmetry between Randy and Peter, either by arguing that Peter is irrational or by arguing that Randy is rational. The former seems very implausible, for most ordinary speakers as well as many trained semanticists

³⁷ We will see shortly (sect. 3.3) how NPM could try to deny this.

would then count as irrational³⁸: all these subjects are enlightened, and yet they take (3)-(4) to differ in truth value. Moreover, as Braun remarks, ordinary speakers like Peter have not been exposed to the rather esoteric evidence for Millianism (having to do with Kripke's anti-descriptivist arguments, the behavior of indexicals, data about quantifying-in and so on). Therefore, they have no reasons to freely substitute proper names within belief contexts as required by Millianism. But Peter cannot be deemed irrational for his not judging (3)-(4) in the way prescribed by a semantic theory he has no reason to believe.

Maybe NPM could grant that Peter is rational but deny that Randy is irrational³⁹. The reason why Randy seems irrational is that (2) is a simple logical consequence of sentences (1)-(9) and it can be easily derived from them by substituting coreferential terms. But a speaker can fail to draw a conclusion (S) which is a simple logical consequence of a set of sentences (Σ) and yet be rational. For instance, one could have a sophisticated semantic theory according to which (S) doesn't follow from (Σ). Or, one might be unable to see that (S) follows from (Σ), like most first-year logic students are unable to see that the indicative conditional ($A \rightarrow B$) follows from (B)⁴⁰.

This move would not eliminate the problematic asymmetry:

- First, Randy is *ex hypothesi* a standard speaker who doesn't hold any peculiar semantic theory about simple sentences. So, unlike the semanticist described above, he doesn't have any sophisticated theoretical reasons to resist substitution. He is also unlike the first-year logic student, for competent reflective speakers don't need any training in logic to see that a sentence like (2) follows from (1)-(9) by simple substitution. So we have no particular reasons to take Randy to be rational.
- Even if Braun could show that Randy is rational, that would not be enough to eliminate the asymmetry. What has to be shown is that Randy is *as rational* as Peter (or, alternatively, that Peter is as irrational as Randy). But this would be wildly implausible. After careful reflection, Randy fails to derive (2) from (1)-(9) *plus* certain logical principles that he routinely employs in his everyday reasoning (Leibniz's law). He has no sophisticated theoretical reasons to do this, nor does he lack the required training. Certainly, he is more irrational/less rational than Peter, who is simply failing to follow a complicated semantic theory he has no evidence for?

³⁸ Braun agrees that speakers who resist substitution about reports are rational, although he claims they might not be *fully* rational; v. Braun (1998, p. 589).

³⁹ Braun and Saul (2002, p. 22) seem to deny that speakers who resist substitution about simple sentences are irrational.

⁴⁰ This argument is offered by Braun (1998, pp. 586-587), although he uses it to argue for a different claim.

It seems hard to deny that there is *some* difference in rationality between Peter and Randy which the Millian has to explain⁴¹. I will now discuss a way in which NPM could try to offer such an explanation.

3.3. Second Reply: Anti-Millian Principles

There is a way for NPM to deny (C). This would block the Asymmetry objection and save a psychological explanation based on (Sch). An argument in this direction is suggested by Salmon in his reply to Schiffer (1987, 2006)⁴². When it comes to specifying what “suitably different ways” amount to, Salmon would endorse option B but reject (C): only Peter, *but not Randy*, has suitably different ways of taking the relevant proposition. Moreover, Salmon would argue that this is true even though, being enlightened, *both* Peter and Randy lack suitably different ways of thinking about the *individual* Superman/Clark.

Suppose Peter has certain views about the semantics of reports like (3)-(4), such that he takes them to express *different propositions* even though he knows that “Superman” and “Clark Kent” are coreferential. According to option B, Peter will then have suitably different ways of believing/disbelieving the proposition expressed by (3)-(4): he takes one and the same proposition, the one expressed by (3)-(4), to be two distinct propositions. However, he doesn’t have such suitably different ways in virtue of his being unenlightened about Superman/Clark, but in virtue of his beliefs about the semantics of reports. Therefore, it doesn’t follow that Randy *also* has suitably different ways of taking the proposition expressed by (1)-(2), for Randy might not have analogous beliefs about the semantics of simple sentences. So (C) is false. Therefore, the NPM theorist would conclude borrowing Salmon’s argument, we *can* explain the difference in rationality between Peter and Randy by appealing to ways of believing. Because of his beliefs about the semantics of reports, Peter has suitably different ways of taking the proposition expressed by (3)-(4) and is rational in resisting substitution. Since Randy has no such beliefs about simple sentences, he doesn’t have suitably different ways for (1)-(2) and his resistance to substitution is irrational.

I grant that, if he helps himself to the solution above, the Millian can account for the rationality of enlightened speakers by appealing to his metaphysics of ways; to this extent, Braun is right. I deny, however, that the resulting account will *exclusively* rely on the standard Millian metaphysics and be purely *psychological*.

According to the response above, Peter has suitably different ways *because* he endorses something like the following principle, which is of course inconsistent with Millianism:

⁴¹ Braun (2006, p. 378) and Salmon (2006) seem to agree.

⁴² Cf. Salmon (1989, pp. 264-272) and Salmon (2006, pp. 372-374).

- 10) The sentences “Lois Lane believes that Superman can fly” and “Lois Lane believes that Clark Kent can fly” express different propositions⁴³.

Crucially, this shows that the explanation offered by NPM doesn’t *exclusively* rely on ways of believing. It’s Peter’s acceptance of the anti-Millian principle (10) which explains his having suitably different ways; but then, since NPM holds that Peter’s having suitably different ways is what explains why he is rational, (10) will be what *really* accounts for Peter’s rationality. The fact that he has suitably different ways of taking the relevant proposition is only a consequence of his anti-Millian credences. NPM can explain speakers’ rationality by appealing to ways, but only through a prior appeal to anti-Millian principles. Ways alone are not enough.

Invoking anti-Millian principles like (10) thus simply creates a new *explanandum* for the non-pragmatic Millian: instead of having to explain why speakers are rational in their anti-substitution intuitions, he must now explain why they accept (10). And, of course, this explanation must not rely on the semantics or pragmatics of attitude reports, or it will not be entirely psychological and NPM’s main thesis will be false⁴⁴.

In what follows, I will argue that NPM’s appeal to (10) is problematic in itself, and that even if it was acceptable, NPM would still be unable to explain why speakers accept (10). I will also argue that the only way for NPM to deal with these worries would be to appeal to pragmatic factors. If true, this shows that the non-pragmatic

⁴³ Alternatively: “The sentences ‘Lois Lane believes that Superman can fly’ and ‘Lois Lane believes that Clark Kent can fly’ express different *thoughts*/have different *contents*/...”.

⁴⁴ Braun usually employs a *different* anti-Millian principle to account for the rationality of enlightened speakers like Peter (Braun 1998, p. 582; Braun 2006, pp. 377-378):

- 11) If Lois Lane is rational, reflective, and attentive, and she believes that Clark Kent cannot fly, then she doesn’t also believe that Clark Kent can fly.

Peter’s reasoning is supposed to go thus: “Lois believes that Clark cannot fly, since she dissents from (2); assuming (11), it follows that she doesn’t *also* believe that Clark can fly; therefore, (4) is false even though (3) and (9) are true”. Accepting (11) (which is again inconsistent with Millianism) gives Peter excellent *reasons* to think (4) is false, so he is not irrational in resisting substitution with (3)-(4).

The problem with (11) is that, unlike (10), it doesn’t enable Peter to have suitably different ways of taking the proposition expressed by (3)-(4). (Indeed, notice that now Braun’s account doesn’t *at all* rely on the ways in which Peter takes that proposition, but only on the *reasons* he has to resist substitution). So (C) still holds: under both options A and B, if Peter has suitably different ways, Randy also does. Therefore, Braun cannot explain Peter’s rationality by appealing to (Sch), for that would also entail that Randy is rational.

Couldn’t Braun simply appeal to (11) and drop his original psychological account? First, this would destroy one of NPM’s main claims, i.e. that we can explain anti-substitution intuitions about simple sentences in the same way we explain intuitions about reports. Such explanation would have to be based on (Sch), while the explanation based on (11) is not. Second, I hope my arguments in this section will show that NPM’s appeal to *any* anti-Millian principle is bound to fail; if so, then appealing to (11) rather than (10) would not help the non-pragmatic Millian.

response to the Asymmetry objection fails, and that the Millian will be forced to adopt a pragmatic account if he wants to explain the rationality of enlightened speakers⁴⁵.

If speakers adopt the semantic principle (10), it is presumably in virtue of their holding some further anti-Millian view about the semantics of reports. For instance, they might be “closet Fregeans” and think that “believes” expresses a two-place relation between a subject and a Fregean proposition, whose constituents are the descriptive senses expressed by proper names within attitude contexts⁴⁶. Alternatively, they might think that “believes” expresses a three-place relation between a subject, a singular non-descriptive proposition and a contextually supplied way of believing that proposition⁴⁷. Other options are open, but one thing is clear: if speakers believe (10), they do so in virtue of their (more or less implicitly) believing one of the semantic theories which are opposed to Millianism.

Clearly, this has several problematic consequences (it’s certainly not an accident that Braun’s original account tried to make do without anti-Millian principles). First, we are now assuming that speakers adopt a wrong semantic theory for attitude reports, while they adopt no such misleading theory for simple sentences. But it would be bizarre if this kind of error was *unique* to reports. In order for the appeal to anti-Millian principles to be plausible, the NPM theorist must find other types of sentences for which speakers endorse a false semantic theory which leads them to have systematically mistaken intuitions. Of course, a Millian could respond by trying to explain our mistaken views *pragmatically*. Semantics-pragmatics confusions are widespread and arise for all kinds of sentences, both simple and not: if our endorsement of (10) results from such a confusion, the Millian will not have to postulate a *special* source of error for reports. Clearly, however, this avenue of response is blocked for NPM, which must therefore find some alternative solution to the problem.

One possibility would be to explain mistaken views about reports as the result of mistaken views about simple sentences. According to this line of argument, speakers take (3)-(4) to express different propositions because they *also* take (1)-(2) to express different propositions⁴⁸. Presumably, Peter takes (3) to say that Lois believes the proposition expressed by (1); *mutatis mutandis* for (4) and (2). Therefore, the NPM theorist might claim, if Peter believes that (1)-(2) express different propositions he will infer that (3)-(4) also do. The problem with this explanation is that it will again entail that Randy is rational. If ordinary speakers take the simple sentences (1)-(2) to express different propositions, Randy also does; but then he will have suitably different ways of

⁴⁵ Schiffer (2006) briefly considers the possibility of appealing to anti-Millian principles to save Millianism. His objections are different from mine (although I do share some of his concerns) and don’t seem very strong to me, but I won’t discuss this here.

⁴⁶ Salmon takes ordinary speakers to be “closet Fregeans” in his (1989, p. 268) and (more explicitly) in his (2006, pp. 372-374).

⁴⁷ Cf. Crimmins and Perry (1989), Richard (1990). Such views will be extensively discussed in ch. 3.

⁴⁸ This is suggested by Salmon (2006, p. 373).

taking the proposition expressed by (1)-(2) and he will count as rational according to (Sch). NPM must therefore postulate a set of anti-Millian principles *specifically targeted on reports* and not on simple sentences, which gives rise to the worries raised in the last paragraph.

A related problem is that, by appealing to (10), the non-pragmatic Millian is explaining away contrary intuitions in the following way: “speakers have anti-Millian intuitions because they believe a semantic theory which is inconsistent with Millianism”. By itself, this is clearly not an acceptable response; *any* semantics could be saved this way, no matter how large the amount of contrary intuitive data. A satisfactory error theory will also have to explain *why* speakers would ever endorse the false semantic view in the first place. Moreover, the appeal to anti-Millian principles ascribes a very serious form of “semantic blindness” to ordinary speakers, thus weakening the very evidence *in favor* of Millianism. If speakers are systematically misled by false principles in their intuitions, why take seriously pro-Millian data like anti-descriptivist intuitions about proper names? Again, notice that a pragmatic account seems to be the only way for the Millian to answer this worry. PM postulates a much less serious form of semantic blindness than NPM: according to PM, speakers are right about *something* in their anti-substitution intuitions, since they correctly take the pragmatically conveyed propositions to differ in truth value. Their only mistake is to confuse such propositions with the semantic content of the reports themselves.

Even if NPM could answer the worries above, it would still have to offer a psychological explanation of our acceptance of (10). Nothing in Braun’s picture suggests what that explanation might look like, and there are principled reasons to doubt that NPM will be able to offer the required explanation without appealing to pragmatic factors at any point.

If NPM is true, there is a psychological mechanism “(M)” which causes us to accept (10). But what could (M) be? Presumably, it will be unlike the mechanisms which lead us to make mistakes in cases like the “Moses Illusion” (“How many animals did *Moses* take on the ark?”). Anti-substitution intuitions about reports are not the result of poor concentration or limited cognitive resources; competent, reflective speakers hold on to their judgments about (3)-(4) even after we ask them to pay attention to the relevant identity facts. Given the robustness of anti-substitution intuitions, they cannot just be caused by a “performance limitation” of some sort.

Perhaps more plausibly, one might identify (M) with a set of principles of our *folk psychology*, i.e. our folk theory of other people’s mental states. There is strong consensus in contemporary cognitive science about the existence of “folk theories”: a folk physics, a folk biology and so on⁴⁹. Such theories are generally taken to be “implicitly” or “subpersonally” believed by most humans, to be (relatively) modular and innate, and to become active at certain developmental stages. (M) could then be

⁴⁹ V. Prinz (2002, ch. 8) for a useful overview of the literature.

identified with a set of deep beliefs which are part of our folk psychology and lead us to accept (10). After all, (3)-(4) concern what Lois believes, so some folk-psychological principle might well bear on our evaluation of such sentences. Moreover, this account would be independently motivated (we have strong evidence for the existence of folk psychology) and it would account for the robustness of anti-substitution intuitions (folk theories are deeply ingrained, so the judgments they produce can be expected to be robust).

Notice, though, that (10) is a *semantic* principle: it specifically concerns the content of belief reports containing coreferential terms, and we accept it because of our having certain further semantic views (e.g. because of our closet-Fregeanism; *v. supra*). So it would be implausible to claim that folk-psychological principles can *fully* explain (10)'s acceptance: folk psychology is presumably silent about the content of proper names within attitude contexts and similar issues in the semantics of natural language. Moreover, folk-psychological principles are generally taken to be innate; but, again, it seems odd to claim that speakers reliably become closet-Fregeans at some ontogenetic stage, and that their doing so is not in any way influenced by the semantic or pragmatic features of reports. Presumably, our interactions with other speakers play a crucial role in our coming to endorse (10). For that reason, it seems that a full account of our acceptance of (10) must at some point inevitably appeal to the semantic or pragmatic profile of ascriptions.

To illustrate, here is how a semantic account might go: speakers encode certain information in the semantic content of (3)-(4), so that they come to express different propositions; we are sensitive to the content of other speakers' utterances; therefore, we correctly conclude that the two reports express different propositions, thus coming to accept (10). Since this semantic explanation would be straightforwardly inconsistent with Millianism, a Millian could instead provide a pragmatic story along the following lines: speakers pragmatically convey different propositions by uttering (3)-(4); again, we are sensitive to the information conveyed by other speakers' utterances and realize that different propositions are being transmitted by (3)-(4); however, we mistake such propositions for the contents of the reports themselves, thus endorsing the false principle (10).

Unlike NPM's appeal to folk-psychological principles, then, a semantic or pragmatic explanation would account for the role that interactions with other speakers must play in our coming to accept (10). This means that, since a semantic explanation is not an option, any Millian who takes ordinary speakers to endorse principles like (10) will have to assume that (3)-(4) pragmatically convey different propositions. Far from making the appeal to pragmatics useless, then, Braun's attempt to explain speakers' rationality by appeal to ways of believing *requires* adopting a PM account! I conclude that the non-pragmatic Millian won't be able to explain in a purely psychological way why speakers accept (10): but this constitutes sufficient reason to reject NPM, which

needs a psychological explanation of (10)'s acceptance in order to offer a plausible account of enlightened speakers' rationality⁵⁰.

In conclusion: Braun is right to claim that the Millian can explain the rationality of anti-substitution intuitions about simple sentences and reports in the same way, i.e. by appealing to his metaphysics of belief. However, the Asymmetry objection shows that such explanation will have to rely on (10) to explain the rationality of enlightened speakers like Peter. So NPM must offer a purely psychological, non-semantic and non-pragmatic account of our acceptance of (10). But I have argued, first, that NPM's appeal to (10) is problematic in itself and, second, that we cannot offer a psychological account of its acceptance. The only way for the Millian to explain why ordinary speakers endorse (10) would require assuming that (3)-(4) pragmatically convey different propositions, in which case NPM's account would not be non-pragmatic anymore. This has two important consequences:

- The non-pragmatic Millian cannot explain in a purely psychological way why ordinary speakers are rational in their anti-substitution intuitions, so NPM fails.
- Since the Millian can *only* explain those intuitions by offering a pragmatic account, if the standard objections against PM succeed then Millianism must be rejected altogether.

I will discuss some of the objections against PM in section 5 *infra*; before doing that, however, I would like to examine another important aspect of Braun's non-pragmatic strategy.

⁵⁰ Isn't the pragmatic strategy just outlined going to be psychological after all? The NPM theorist might stress how psychological facts play a crucial role in all pragmatic explanations (cf. for instance classic Gricean accounts or Relevance Theory), so that even PM accounts will count as psychological.

Of course, I grant that, on this reading of "psychological", the Millian *can* offer a psychological explanation of speakers' rationality. But this is irrelevant for my purposes; what I deny is that the Millian can provide an explanation which "does not assume that utterances of (3) and (4) pragmatically convey different propositions" (Braun 1998, p. 579). This is the psychological account I am attacking, and this is what Braun himself means by "psychological explanation". Notice that, by Braun's own definition, if an explanation assumes *at any point* (i.e. even when explaining (10)'s acceptance) that (3)-(4) pragmatically convey different propositions, that explanation will still count as non-psychological.

4. NPM And Behavior

There is a second “classic” objection that all Millians have to face: if Millianism was true, we could not account for the role that certain intentional generalizations seem to play in the explanation and prediction of behavior⁵¹. Braun (2000, 2001a,b) has tried to offer a non-pragmatic response to the objection; I will argue that his response fails, and that a Millian will again be forced to appeal to what intentional generalizations allegedly convey at the pragmatic level in order to account for their role in explanation/prediction.

Consider the following generalization⁵²:

- 12) If a person wants Twain to autograph her book, and she believes that if she waves then Twain will autograph her book, then, other things being equal, she will wave.

Suppose Lucy is not aware of the relevant identity claim (“Twain is Clemens”) and wants Twain to autograph her book. Also, suppose she sincerely says: “If I wave then Clemens will autograph my book”. Presumably, Lucy believes that if she waves then Clemens will autograph her book. But then, *if NPM is true*⁵³, Lucy also believes that if she waves then *Twain* will autograph her book. So if NPM is true Lucy satisfies the antecedent of (12). However, in the circumstances described Lucy might still *not wave* (she might reason: “If I wave then Clemens will autograph my book; but I want Twain, and not Clemens, to autograph my book”). So there are possible cases where an agent satisfies the antecedent of (12) but not its consequent. Therefore, (12) has possible counterexamples and is a false generalization. But (12) is true, so NPM is false. The problem is particularly serious, since it seems we can use (12) to explain/predict the waving behavior of someone who wants Twain to autograph her book and believes that if she waves then Twain will autograph her book. But if (12) was false we could not use it to produce sound explanations and predictions!

According to the objection, Lucy’s not waving will constitute a counterexample to (12) once we assume NPM. In reply, defenders of NPM⁵⁴ have employed what I will call “the *ceteris paribus* strategy”: even assuming NPM, Lucy’s behavior is not a counterexample to (12) but rather a tolerable exception to it, since other things are *not* equal in her case. If NPM is true, Lucy both wants Twain to autograph her book and

⁵¹ V. Richard (1990, 1997).

⁵² I switch to a Twain/Clemens case to follow Braun’s examples.

⁵³ I present the objections in this section as aimed against NPM only, for PM would have a different response against them; v. *infra*.

⁵⁴ V. Braun (2000, 2001a,b). The same strategy is also used by Schneider (2005, 2011), although for slightly different purposes (v. ch. 3, sect. 4.2). The move was originally proposed by Fodor (1994).

believes that if she waves then Twain will autograph her book. So she does satisfy the first two conjuncts of (12)'s antecedent. However, because she is unenlightened, she believes and desires those two propositions in “mismatching” or “suitably different” ways (v. previous sections); the mental sentences in her belief and desire box don't “functionally interact” in the same way they would in an enlightened subject, so she doesn't wave. Given the mismatch in her ways of believing/desiring, however, other things are *not* equal in her case and she constitutes a tolerable exception rather than a counterexample to (12).

It might be objected that the move seems entirely *ad hoc*: why should we take (12)'s *ceteris paribus* clause to include a “matching ways” condition, so that other things are not equal with unenlightened subjects? In reply, proponents of the strategy have attempted to provide independent reasons to treat subjects like Lucy as tolerable exceptions⁵⁵. As I will now argue, however, (12) is not the *crux*: even if NPM could offer independent reasons to include a “matching ways” condition in the *ceteris paribus* clause of (12), it would *still* be unable to account for a large class of generalizations which appear just as explanatory and predictive as (12). So the *ceteris paribus* move, even if independently motivated, would still not enable NPM to account for our actual explanatory practices.

Consider the following generalization:

- 13) If a person wants Twain to autograph her book, believes that if she waves then Clemens will autograph her book, *and doesn't believe that Twain is Clemens*, then, other things being equal, she will *not* wave.

Like (12), (13) seems true; like (12), it seems we can use it to predict and explain behavior. The difference with (12), of course, is that (13) can be used to predict/explain a subject's *not* waving rather than her waving. Indeed, (13) seems to be precisely the generalization we would use to explain why Lucy doesn't wave in the circumstances described above. Notice, though, that Lucy (as well as any other intentional subject) will be *unable* to satisfy (13)'s antecedent once we assume NPM. Everyone believes that Twain is Twain; if NPM is true, this entails that everyone believes that Twain is Clemens; therefore, under NPM it is impossible for an intentional subject like Lucy to satisfy the antecedent of (13). But then (13) cannot be used to explain/predict Lucy's not waving, since she doesn't satisfy the antecedent. Indeed, if NPM is true (13) cannot be used to explain/predict the behavior of *any* intentional subject, for no one will ever satisfy its antecedent; but (13) can be so used, so NPM is false⁵⁶.

⁵⁵ V. especially Fodor (1994) and Schneider (2005, 2011).

⁵⁶ Notice that the *ceteris paribus* strategy would not be of any use here. That strategy can at best show that Lucy is a tolerable exception rather than a counterexample to (12). The problem with (13), however, is *not* that Lucy is a potential counterexample to it, but that she cannot satisfy its antecedent once we assume NPM. There is no way for NPM to deny this by appealing to the *ceteris paribus* clause of (13).

Contrast NPM with any non-Millian semantics “NM” on which substitution of coreferential terms is blocked within propositional attitude contexts. NM can easily account for the explanatory/predictive power of *both* (12) and (13). Assuming NM, Lucy doesn’t satisfy the antecedent of (12); she does want Twain to autograph her book, but she only believes that if she waves then *Clemens* will autograph her book. So NM has no problems with (12), and without having to appeal to the *ceteris paribus* strategy. NM also has no problems with (13). Assuming NM, Lucy (as well as any other intentional subject) can perfectly well satisfy (13)’s antecedent; she does believe that Twain is Twain, but she doesn’t thereby *also* believe that Twain is Clemens. So under NM, but *not* under NPM, (13) can be used to explain/predict the behavior of people like Lucy⁵⁷.

The problem, in sum, is that the *ceteris paribus* strategy shows *at best* that NPM is *consistent* with generalizations like (12). Even then, however, it will still follow from NPM that generalizations like (13) cannot be used to explain/predict behavior⁵⁸. It now seems rather irrelevant whether NPM can appeal to (12)’s *ceteris paribus* clause to avoid the first objection. Even if we grant that move, a large class of generalizations which seem just as explanatory and predictive as (12) will still be left out as useless for purposes of explanation/prediction! (On the contrary, *both* kinds of intentional generalizations will be straightforwardly captured by NM). The heated debate⁵⁹ about whether to include a “matching ways” condition in (12)’s *ceteris paribus* clause appears a lot less urgent once we realize that, in any case, there is a large class of intentional generalizations that NPM has no hope to capture.

Is there any way for the Millian to account for our use of (13)? The only available option, again, would be to adopt a PM account. While a pragmatic Millian would also have to grant that we cannot “use” the semantic content of (13) for explanation and prediction, he could still appeal to what an utterance of (13) will allegedly convey. For instance, following Salmon’s “metalinguistic” proposal (v. sect. 2.3), he might hold that the utterance will convey a proposition having *<... and doesn’t believe that “Twain is Clemens” is true>* as the third conjunct of the antecedent. The antecedent of this proposition can clearly be satisfied by intentional subjects like Lucy, even assuming Millianism. The pragmatic Millian might then hold that, in a broad sense of “using” which includes what is pragmatically conveyed by a generalization, we *can* use (13) to explain/predict behavior. Since no such move is available on Braun’s NPM, the only option left for the Millian who wants to account for (13) will against consist in appealing to pragmatics.

A possible reply in defense of NPM is suggested by Schneider (2005, 2011), who would argue that NPM *can* explain/predict Lucy’s not waving. As we have seen, NPM holds that, if Lucy believes and desires the relevant propositions in mismatching ways,

⁵⁷ We will go back to non-Millian views at length in ch. 3.

⁵⁸ I will say more about intentional explanation and prediction in ch. 2 (sect. 3.1).

⁵⁹ V. especially Arjo (1996) and Aydede and Robbins (2001), who respond to Fodor (1994).

then she will not wave. So her behavior can be predicted and explained at the “computational” level of the theory: if the LoT sentences in Lucy’s belief and desire box are disconnected in the right way (due to her being unenlightened), then she will not wave. So NPM is perfectly able to explain/predict Lucy’s behavior, although admittedly it cannot use (13) to do so⁶⁰.

This reply would miss the point. I agree that the NPM theorist can predict/explain Lucy’s behavior by appealing to something like the following generalization:

- 14) If a person wants Twain to autograph her book, believes that if she waves then Clemens will autograph her book, *and wants and believes these propositions in mismatching ways*, then, other things being equal, she will not wave.

I grant that (14) is as explanatory and predictive as (13), and that Lucy can perfectly well satisfy its antecedent even assuming NPM. Still, these considerations fail to address my objection. NPM is a *semantic* theory; it is a theory about what is semantically expressed and pragmatically conveyed by our reports of propositional attitudes. If that theory was true, (13) could not be used to predict/explain Lucy’s behavior; but since (13) can be so used, that semantic theory is false. The fact that we can *also* use (14) to predict/explain Lucy’s not waving cannot save NPM as a theory of ascriptions, for NPM is still unable to account for the generalizations *we actually use*, i.e. (13).

Is there some other response available to the non-pragmatic Millian? There is no way to deny that, if NPM is true, then (13) cannot be used to explain/predict behavior; therefore, NPM can only deny that (13) can be so used and explain away our intuitions to the contrary as incorrect⁶¹. The story will presumably go as follows. We think that (13) can be used to explain/predict Lucy’s behavior because we think she can fail to believe that Twain is Clemens, thus satisfying (13)’s antecedent. This intuition is wrong, just like our parallel intuition about (4): Lois *does* believe that Clark can fly, and Lucy *cannot* fail to believe that Twain is Clemens. Presumably, our intuitions about (13) will then be explained as the product of anti-Millian principles like (10), following the strategy outlined in sect. 3.

Clearly, this concedes a lot to NPM’s opponents. Granting that (13) cannot be used to explain/predict behavior is a very substantial cost, *especially* if one agrees that (12) can be so used, as Braun does; *prima facie*, the two generalizations seem exactly alike in their explanatory/predictive power. Moreover, we have seen that NPM has to explain away our intuitions about (13) by appealing to anti-Millian principles like (10). But this means that, if my previous arguments against NPM’s appeal to (10) succeed, the non-pragmatic Millian will simply have no way to explain those intuitions.

⁶⁰ I will examine Schneider’s view in detail in ch. 3 (sect. 4.2).

⁶¹ Braun does consider a similar move (with respect to (12)) as a possible alternative to the *ceteris paribus* strategy: v. e.g. Braun 2000 (sect. 11).

Given the substantial costs of this line of response, it is not surprising that, when answering the original objection, Braun tried to capture our intuitions about (12) rather than explain them away as incorrect. What my argument shows is that, when it comes to (13), this possibility is not available; there is no way for the non-pragmatic Millian to capture the role that such generalizations intuitively seem to play in our everyday explanations and predictions of behavior.

5. Some Problems With Pragmatic Millianism

We have considered two classic objections against Millianism: the view has troubles explaining the rationality of ordinary speakers and accounting for the role of intentional generalizations in the explanation and prediction of behavior. I have discussed Braun's attempt to offer non-pragmatic solutions to both problems, and argued that neither of them works. If there is any hope for the Millian to deal with these issues, it lies in adopting a pragmatic account. Unfortunately, however, such accounts also have to face a number of serious objections, which is precisely what pushed Braun to develop his non-pragmatic alternative in the first place.

A first set of objections has targeted the specific PM accounts proposed by Salmon and Soames (v. sect. 2.3). Soames explains anti-substitution intuitions by claiming that speakers use reports like (3)-(4) to pragmatically assert “descriptively enriched” propositions like those expressed by (7)-(8). In reply, Braun and Sider (2006) have argued that this exposes Soames to standard Kripkean arguments against descriptivism⁶². Suppose Gödel is about to give a lecture and the host introduces him as follows: “We are very pleased to have the person who proved the incompleteness of arithmetic with us today. Professor Gödel will speak on logic”. Smith and Jones arrive late at the lecture and miss the host's first sentence. However, they do hear the host utter “Professor Gödel will speak on logic”. In addition, they know of Gödel's theft and mistakenly think everyone else does, too. Smith then says to Jones: “How can Gödel have the nerve to speak on logic after having stolen the incompleteness proof from Schmidt? Still, the host believes that Gödel will speak on logic. So perhaps he will”. Now, as Braun and Sider point out:

According to Soames, by uttering ‘The host believes that Professor Gödel will speak on logic’, Smith primarily asserts the descriptively enriched proposition **The host believes that Professor Gödel, who stole the incompleteness proof from Schmidt, will speak on logic**. Since the host believes no such thing, this proposition is false. Yet [...] our intuition is that Smith's utterance is true [...] (intuitively, Smith asserts *nothing* false) (Braun and Sider 2006, p. 673).

If (3)-(4) are standardly used to assert descriptively enriched propositions, then presumably Smith's utterance is *also* used to assert a descriptive proposition about Gödel. If the proposition in bold is the one asserted, however, Soames' account will mistakenly predict that Smith has asserted something *false* with his utterance.

Similar problems affect Salmon's proposal, on which utterances of (3)-(4) are routinely used to convey the “metalinguistic” (5)-(6). As Saul (1998) points out, this cannot explain anti-substitution intuitions in all cases. Suppose Lois is only acquainted with Superman/Clark “demonstratively”: she ignores the name of both the mighty superhero she has seen flying and of the reporter with glasses she has seen in the office.

⁶² They also raise a similar objection against Soames' theory of simple sentences, but I prefer to focus on reports here.

Clearly, Lois would accept *neither* “Superman can fly” *nor* “Clark Kent can fly”. Now, an ordinary speaker who knew all this would *still* think (3) was true and (4) false; however, he would now take *both* (5) and (6) to be false. Therefore, we cannot explain his anti-substitution intuitions about (3)-(4) as the result of his intuitions about (5)-(6).

Salmon’s proposal could be patched by construing the “guises” mentioned in (5)-(6) as something *other* than natural language sentences: we could for instance identify them with mental representations (e.g. LoT sentences), as on Braun’s construal of ways of believing. While this might avoid the objection against the original meta-linguistic proposal, however, Braun notes that *any* appeal to “guises” in (5)-(6) will still risk being implausible for a different reason⁶³. Ordinary speakers have no familiarity with notions like “BEL-relation” and “propositional guise”. Therefore, they would not understand (5)-(6), nor is there any context in which they ever utter these sentences. But how could speakers’ utterances of (3)-(4) systematically convey propositions that the speakers themselves never express or even understand? (Of course, as Braun notes, the worry will be even stronger once we adopt a sophisticated construal of guises as LoT sentences⁶⁴).

Besides the problems affecting specific PM accounts, there is a more general issue that all pragmatic Millians have to face. As Braun (1998, pp. 570-571) notes, if utterances of (3)-(4) had the same semantic content while conveying different propositions, speakers should be able to recognize this and consequently retract their original anti-substitution intuitions. For instance, Salmon (1989, pp. 252-53) regards the implicatures triggered by belief reports as analogous to the generalized implicatures produced by sentences like “Jane became pregnant and she got married”. Ordinary speakers usually think that the temporal ordering of the events in question is relevant for the truth value of these sentences. However, after appropriate “training” they come to realize that such sentences only “suggest” something about temporal order. But ordinary speakers will *still* take (3)-(4) to differ in truth value after we bring up the possibility that they might only differ in what they “suggest”. More importantly, trained semanticists who are usually perfectly capable of recognizing the distinction between the semantic and pragmatic content of an utterance will often have the same response. Anti-substitution intuitions are just too robust to be explained in a purely pragmatic way.

I am persuaded by some of the standard objections against PM. In particular, I agree with Braun that anti-substitution intuitions are too widespread and robust to be the outcome of a mere semantics/pragmatics confusion⁶⁵. Now, I have argued in previous

⁶³ Braun (1998, p. 567).

⁶⁴ The revised version of Salmon’s proposal might also be vulnerable to other arguments in Saul (1998), but I cannot discuss this here.

⁶⁵ On the other hand, some of the *specific* objections against Salmon and Soames’ accounts could *also* be used against the non-Millian, “contextualist” account I favor (v. ch. 3 *infra*): cf. in particular the

sections that non-pragmatic versions of Millianism fail: if Millianism is true, then pragmatic Millianism is. Therefore, I take the objections against PM to show that Millianism must be rejected *in toto* as a theory of ascriptions. Clearly, this is not the end of the story. There are various moves that pragmatic Millians could attempt in response to those objections, and further worries that could be raised against PM⁶⁶. However, I will not discuss the issue further, for two reasons. First, a complete analysis of PM and its problems would require a dissertation of its own. Second, and more importantly, my arguments in the following chapters would *also* go through (with some modifications) if we assumed pragmatic Millianism instead of a non-Millian semantics (I will explain why in ch. 3, sect. 3). So, while I will take anti-PM arguments at face value and assume that the right view of ascriptions must be non-Millian, the substance of my theory of concepts would be left untouched if some pragmatic Millian account could be made to work after all.

The failure of Millianism is the first important moral that a successful theory of reports must take into account. There is, however, a second aspect of our attitude-reporting practices that should play an equally important role in that theory, and that will also be crucial for our later discussion: the truth-conditions of attitude reports seem to *vary with context*. We will develop this point extensively in chapter 3: to anticipate, here is an interesting case by Jennifer Saul:

Suppose I am discussing what people in general think of Bob Dylan's singing abilities, and the person I'm talking to knows him only as 'Bob Dylan'. I've been told (truthfully) that Glenda, a childhood friend, who knows him only as 'Robert Zimmerman', believes that he has a beautiful voice. Specifically, someone I trust has uttered sentence (1):

(1) Glenda believes that Robert Zimmerman has a beautiful voice.

I may report this with sentence (2):

(2) Glenda believes that Bob Dylan has a beautiful voice.

(2) seems true, even though Glenda would never assent to it. Further, we have no hesitation in saying this, although we know next to nothing about how Glenda may think of Bob Dylan/Robert Zimmerman. [...] Suppose now that Glenda is participating in a marketing poll which asks for her opinions of various singers' voices, by name. One of the names on the list is 'Bob Dylan.' I'm asked to predict her responses. It would be wrong for me to reply with 'Glenda believes that Bob Dylan has a beautiful voice', even though that very sentence seemed true in our previous context. This tells us quite straightforwardly that the truth conditions of attitude-reporting sentences seem to vary with context (Saul 1999a, pp. 358-59).

arguments by Saul (1998) and Braun and Sider (2006). I believe those objections could be overcome by the contextualist, but I won't be able to discuss this here.

⁶⁶ V. Soames (2006 a,b) for discussion.

When conjoined with the results of our anti-Millian arguments, Saul's considerations seem to show that a successful theory of ascriptions must both take into account their context-sensitivity and be non-Millian. In chapter 3, we will examine a family of theories that were specifically designed to satisfy these two *desiderata*. As we'll see, such theories don't just constitute a promising solution to the problem of ascriptions: perhaps more surprisingly, they might also shed light on the nature of concepts itself. To see why, however, we must first say a bit more about some recent debates on concept individuation and possession; this will be our starting point in the next chapter.

Chapter 2

Publicity Vs. Holism

1. Introduction

This chapter has two main goals. The first one is to provide a basic framework for our discussion of concept individuation and possession. That discussion will take place against the background of a specific view, the so-called “Representational Theory of The Mind” (RTM); I describe the basic tenets of RTM in section 2.1. I then turn to examine a principle which many have taken to follow straightforwardly from RTM itself (sect. 2.2). To put it in simple terms, that principle holds that two subjects who are covered by the same intentional generalization must also have certain specific concepts in common. It follows from this principle that concepts are “public” entities which are routinely shared by ordinary thinkers: the principle is thus standardly referred to as the “publicity principle”.

Having summarized the main arguments that have been offered in favor of publicity, I’ll turn to my second goal. I will examine a specific family of theories of concept individuation/possession, namely those “holistic” views according to which concepts are individuated by their global inferential role in a system (sect. 2.3). There is a close connection between holistic theories and the publicity principle: as we will see, many have tried to show holism to be straightforwardly *inconsistent* with publicity. Having summarized these standard anti-holistic arguments, I will discuss two maneuvers that have been attempted to defend holism from publicity-based objections, and argue that they both fail (sect. 3). In section 4 I will consider a possible alternative response, one that has been largely ignored in the literature and that rejects the original argument for publicity as *invalid*. I will suggest that there is no obvious way to patch the argument, and that this has two important consequences. First, the standard anti-holistic arguments examined in section 2.3 are blocked until an alternative argument for publicity is provided. Second, we need to find a new way to answer the question which motivated the argument for publicity in the first place, namely: does the interpersonal applicability of intentional generalizations require concepts to be public? In response to this question, I’ll suggest that there is an alternative way to establish the publicity principle, one which is based on our best semantics for attitude ascriptions: this will lead us to the next chapter, where this alternative route to publicity will be explored in more detail.

2. The Publicity Principle

2.1. The Representational Theory Of The Mind (RTM)

As anticipated, our discussion will take place within a specific theoretical framework. This is the framework against which many contemporary debates about concepts have been conducted, and I will take it for granted throughout most of our discussion. The framework in question is known as the “Representational Theory of the Mind” (RTM). I will characterize RTM as the conjunction of the following claims (my reconstruction is mostly based on Fodor (1998, pp. 7-39) and Aydede (2000a, pp. 23-24):

RTM (I): A subject S has a propositional attitude *pa* with propositional content P just in case S stands in an appropriate functional relation to an internal *mental representation r* expressing proposition P.

RTM (II) A *mental representation r* expressing a proposition P is a structured entity constituted by a *set of concepts* [C₁, C₂, ... C_n].

Let's say a bit more about these two claims⁶⁷:

RTM (I): According to RTM (I), a subject S believes that P just in case S stands in the appropriate belief-relation to an internal mental representation *r* expressing P; a subject S desires that P just in case S stands in the appropriate desire-relation to an internal mental representation *r* expressing P; and so on for all other kinds of propositional attitudes. Following standard usage, I will refer to mental representations which express propositions as “*thoughts*” (so [*there is beer in the fridge*] is a thought, whereas [*cold beer*] is not⁶⁸).

The relations that must obtain between a subject and a thought in order for that subject to have a certain propositional attitude are functionally defined, in the following sense. A mental representation *r* expressing proposition P constitutes a belief that P (rather than, say, a desire that P) just in case *r* plays the *functional role* of a belief. For

⁶⁷ I intend my characterization of RTM to be as broad and ecumenical as possible. In particular, I do *not* mean to identify RTM with the Language of Thought (LoT) hypothesis (cf. Fodor 1975, 2008). Admittedly, the LoT hypothesis is the most popular and fully worked-out version of the RTM program. Still, one might accept RTM (I)-(II) without also identifying mental representations with language-like symbols on which mental processes defined over the symbols' syntactic features operate. Whether alternative versions of RTM are possible is not an issue I can discuss here, but it's at least not obvious that RTM (I)-(II) will force us to endorse the LoT hypothesis. (This also constitutes the main difference between my way of presenting RTM and Fodor (1998)/Aydede (2000a), who both describe the view as a version of the LoT theory).

⁶⁸ In this section, I will refer to mental representations by using expressions in brackets and italics.

instance, if r is a belief that there is beer in the fridge, then r should be such that, when you are in representational state r and you want to have some beer, then you will open the fridge other things being equal (you are not paralyzed, blind and so on). So a subject will stand in the belief-relation to a thought just in case that thought plays the functional role of a belief; *mutatis mutandis* for all other kinds of propositional attitudes. (For brevity, I will say that a subject “accepts” a thought whenever he stands in the belief-relation to that thought⁶⁹).

RTM (II): According to this thesis, concepts are the constituents of mental representations. How should we interpret this claim? The answer will depend on the ontology of concepts we adopt. As anticipated, two options have been especially prominent in the literature: concepts can be identified with abstract objects or mental representations. RTM (II) can be accepted on both of these views, but the claim will be interpreted differently depending on our background ontology:

- According to *Fregean* RTM ontologies⁷⁰, concepts are abstract objects (e.g. Fregean senses) which are *expressed* by mental representations: for instance, my mental representation [dog] expresses the abstract concept DOG⁷¹. On this view, a mental representation r with propositional content P is “constituted” by concepts in a relatively loose sense: concepts are simply the *contents* of the basic mental representations which compose r . For instance, [dogs bark] is a mental representation constructed from the mental representations [dogs] and [bark], which in turn express the abstract concepts DOGS and BARK⁷².
- According to *non-Fregean* RTM ontologies, concepts *are* mental representations⁷³. On this view, my concept DOG is a mental representation referring to dogs. My mental representation [dogs bark] will then be literally constituted by the concepts DOGS and BARK, since these just *are* the basic mental representations from which the structured representation in question is constructed.

⁶⁹ In passing, note that the belief-relation to mental representations is in fact a species of Salmon’s BEL-relation (v. ch. 1).

⁷⁰ V. especially Peacocke (1992, 2005), Rey (1983, 1985), Glock (2009).

⁷¹ As usual, I refer to concepts by using expressions in small capitals.

⁷² Note that a Fregean theorist might want to reserve the label “thoughts” for the *propositions expressed* by mental representations rather than the *mental representations* themselves. This would be incompatible with the definition of “thoughts” I offered above, on which they are identified with mental representations. However, my terminology is relatively standard within RTM and it simplifies the discussion significantly, so I will stick to it in what follows. Nothing turns on this, since Fregean and non-Fregean ontologists agree that concepts are the constituents of thoughts in one of the senses delineated here.

⁷³ V. especially Fodor (1998), Laurence and Margolis (1999, 2007).

Each of these ontological views is adopted by at least some authors who subscribe to RTM, and they will both play an important role in our discussion. In what follows, I will remain neutral on which one we should adopt in developing our theory of concepts. I will therefore try to cast my arguments in terms that would be acceptable for both sides; whenever this is not possible, I will formulate my arguments in two versions, one for each of the ontological views at hand.

2.2. The Argument For Publicity

Given RTM, several arguments can be offered in defense of the claim that concepts are “public” entities which are routinely shared by ordinary thinkers:

- The claim that different subjects can have the same desires, beliefs, intentions etc... seems plausible in and of itself. But, given RTM (I), it seems that two subjects can only have the same belief or desire if they stand in the right functional relation to the same mental representation. For instance, two subjects can only believe that dogs bark if they both stand in the belief-relation to *[dogs bark]*. But, by RTM (II), this mental representation is constituted by the concepts DOGS and BARK, so it seems that two subjects could not both stand in the belief-relation to it unless they both had its constituent concepts DOGS and BARK.
- In order for *communication* to be successful, it must be possible for those who exchange information to come to be in the same mental state. If I tell you that dogs bark, I will have succeeded in communicating the relevant information only once I get you to entertain the thought *[dogs bark]* (which you might then decide to accept or reject). But, again, this thought is constituted by DOGS and BARK, so we could not both entertain it unless we both had its constituent concepts DOGS and BARK⁷⁴.
- Finally, it seems that in order for me and you to *genuinely agree* on a certain subject matter we must be in the same mental state. For instance, we won’t agree on whether dogs bark unless we both believe that dogs bark. But, for the reasons already given, this will not be possible under RTM unless we also have the relevant concepts in common (in this case, DOGS and BARK). *Mutatis mutandis* for genuine *disagreement*, which will only occur when we (respectively) accept and reject one and the same thought⁷⁵.

All these considerations would deserve careful analysis, and we will go back to them later. Throughout chapters 2-3, however, my focus will be on a different argument

⁷⁴ V. Prinz (2002, p. 14), Fodor and Lepore (1992, pp. 8-9).

⁷⁵ V. the argument by Williamson (2007) that was discussed in the main Introduction. V. also Fodor and Lepore (1992, pp. 11-13) and Sainsbury and Tye (2012, p. 21).

that has been offered in defense of concept publicity. Many RTM theorists have argued that concepts must be shared in order for *interpersonally applicable intentional generalizations* to be possible. The main reason why I will focus on this argument is that it has played a more central role in recent disputes about concepts⁷⁶. (At the same time, the psychological generalizations argument does bear some structural analogies to other publicity arguments. For this reason, much of what I'll say about it would also apply to the arguments above, for reasons that will be discussed later).

Several RTM theorists have argued that the interpersonal applicability of intentional generalizations and explanations requires concepts to be public:

Publicity seems to be the backbone of explanation and prediction of human behavior. It is by attributing propositional attitudes that the folk explain and predict people's behavior [...]. It appears that this activity requires that subjects be subsumed under counterfactual supporting content generalizations that are inter- and intra-personally applicable. Consider some examples: "For any person S, if S desires that P and believes that S can bring it about that P, then S will try to bring it about that P"; "Thirsty people who believe they are thirsty tend to seek water" [...]. There are likely to be occasions where these can be used to cover more than one agent (to explain or predict their behavior simultaneously or diachronically), so that the range of the universal quantifier over people may vary from two people picked out in a certain way for the purposes at hand, to entire communities, or even across nations. When they are deployed successfully, these generalizations attribute the same propositional attitudes to agents. The agents' behavior is supposed to be explained/predicted by subsuming them under such generalizations. So propositional attitudes attributed to agents to explain/predict their behavior are required to be type-identical across them so they can be subsumed by such generalizations. The success of intentional explanation and prediction seems to depend on this fact (Aydede 2000a, pp. 5-6).

Concepts must be capable of being shared by different individuals and by one individual at different times. [...] concepts are implicated in intentional explanations of behavior. An intentional explanation of behavior is one that explains what a person does by appeal to her mental states. For example, Mary opened the liquor cabinet because she desired a glass of scotch and believed she could find some there. As this example illustrates, typical intentional explanations make reference to propositional attitudes, and attitudes are composed of concepts. Perhaps the most striking feature of intentional explanations is their apparent generality. A single intentional explanation can subsume many different people. Felix, Hugo and Greta might all open their respective liquor cabinets for precisely the same reason that Mary did. But, actions can be motivated by the same attitudes only if those attitudes are composed of the same concepts. If intentional explanations generalize, concepts must be shareable (Prinz 2002, pp. 14-15).

On the face of it, concepts are the stuff of which psychological claims and explanations are made: generalizations and explanations of, e.g., cognitive development, fallacies in reasoning, vision and language understanding (to take some of the more successful areas of recent psychology), all these *presuppose* concepts as shared constituents of the propositional attitudes the explanations concern. It's not clear how even to *describe* the phenomenon of the Müller-Lyer illusion unless we can presume that people share a concept of *longer than*; or the gambler's fallacy, without them sharing *more likely* (Rey 2009a, p. 2).

⁷⁶ Two examples: the debate between Fodor/Lepore (1992), Fodor (1998) and Block (1998); and the dispute between Aydede (1998, 2000 a,b) and Schneider (2011).

[...] RTM takes for granted the centrality of intentional explanation in any viable cognitive psychology. In the cases of interest, what makes such explanations intentional is that they appeal to covering generalizations about people who believe that such-and-such, or people who desire that so-and-so, or people who intend that this and that, and so on. In consequence, the extent to which an RTM can achieve generality in the explanations it proposes depends on the extent to which mental contents are supposed to be shared⁷⁷. If everybody else's concept WATER is different from mine, then it is literally true that only I have ever wanted a drink of water, and that the intentional generalization 'Thirsty people seek water' applies only to me. [...] Prima facie, it would appear that any very thoroughgoing conceptual relativism would preclude intentional generalizations with any very serious explanatory power [...].

From which Fodor concludes that:

Concepts are *public*: they are the sort of thing that people can, and do, *share* (Fodor 1998, pp. 28-29).

Setting aside some minor differences, all these arguments seem to have the same structure. Before trying to reconstruct that structure in a more perspicuous form, however, we must get clearer on what these arguments are trying to establish. What is it for two subjects to "share a concept" or "have the same concept"? In particular: is it required that the two subjects have two *numerically identical* concepts C_1 and C_2 , or is it only required that they have concepts of the same *kind* or *type*?

Crucially, the answer will depend on the ontology of concepts we choose. Consider the two alternatives described in section 2.1:

- *Fregean RTM ontology*: if concepts are abstract objects which are "expressed" by our mental representations, then two distinct subjects can have two *numerically identical* concepts C_1 and C_2 . For instance, I might have a mental representation [dog₁] expressing concept C_1 , you might have a mental representation [dog₂] expressing concept C_2 , and C_1 and C_2 might be numerically identical with the abstract concept DOG and therefore with each other.
- *Non-Fregean RTM ontology*: if concepts are mental representations, then (by the indiscernibility of identicals) distinct subjects cannot have numerically identical concepts (my mental representation [dog] is mine and not yours). But distinct subjects could still have *type identical* concepts, e.g. two token mental representations [dog₁] and [dog₂] belonging to the same type concept DOG⁷⁸.

In what follows I will leave open whether the publicity constraint, as established by the above arguments, requires strict numerical sameness or mere sameness of type⁷⁹. I will therefore assume that both ontologies can account for the kind of "concept sharing" that publicity requires by adopting one of the two strategies I just sketched.

⁷⁷ Fodor's slipping from "concepts" to "contents" in this passage is significant (v. Glock 2009, pp. 23-29), but in this context we can ignore it and interpret Fodor as referring to concepts alone.

⁷⁸ Cf. Laurence and Margolis (2007, pp. 562-570).

⁷⁹ For discussion, v. Rey (2004), Laurence and Margolis (1999, 2007), Glock (2009).

We can now try to summarize the common structure of the various publicity arguments that were reviewed earlier⁸⁰:

- a) Consider an intentional generalization like:

(G1) If a subject S wants to get water, then other things being equal S will look for water⁸¹.

(G1) seems to “apply” to many subjects. There are several intentional agents who satisfy its antecedent, i.e. who all want to get water. And, if other things are equal (they are not paralyzed, they do not have any overriding desires and so on), these agents will look for water. Such agents are therefore “covered” by (G1), which “applies” to all of them⁸².

- b) Since these agents all satisfy (G1)’s antecedent, they must all satisfy the attitude ascription embedded in it: “S wants to get water” must be true of all of them.
- c) So these agents presumably have a propositional attitude in common, namely the desire to get water. If they didn’t all have this desire, the ascription would not be true of them.
- d) But, according to RTM (I), someone has a propositional attitude *pa* with content P just in case he/she stands in an appropriate functional relation to a mental representation expressing P. In our case, the relevant mental representation is presumably something like [*I get water*], so all the agents to which (G1) applies and who share the desire to get water will stand in the desire-relation to the mental representation in question.
- e) Finally, it follows from RTM (II) that the concept WATER is one of the constituents of the mental representation [*I get water*]. But then all the agents who stand in the desire-relation to [*I get water*] must have the concept WATER. If they didn’t, how could they all stand in the desire-relation to a mental representation that is partially constituted by that concept?
- f) Conclusion: all the agents to which (G1) applies *have the same concept WATER* (equivalently: they “share” WATER or have the concept WATER “in common”).

⁸⁰ For the time being, I prefer to follow the RTM theorists quoted above and state the argument in a relatively informal fashion. We will re-assess the argument more carefully in sect. 4 *infra*.

⁸¹ V. Appendix B for a list of numbered sentences for chs. 2-3.

⁸² In fact, things are a bit more complicated: sometimes a generalization applies to a subject even if that subject doesn’t satisfy its antecedent. We can ignore this complication for now: v. ch. 3, sect. 2.1 (fn. 115) for discussion.

We can now generalize from our example to a broader principle. Let's say that a concept C is "involved" in an intentional generalization when its possession is required for the generalization to apply to someone (like having WATER is required for (G1) to apply to a subject). Then our (a)-(f) argument seems to establish the following principle:

Publicity (PUB): For every intentional generalization G, there is a concept C "involved" in G, such that G applies to a group of subjects only if these subjects all have C.

(Where, remember, we are leaving open whether the subjects to which G applies are required to have the *very same concept* C, i.e. numerically identical concepts, or simply concepts belonging to the same *type* C).

Crucially, note that intentional generalizations apply to the *same subject at different times*, as well as different subjects. For instance, (G1) applies to my current time-slice at time t_1 , as well as my future time-slice at t_2 . So it follows from (PUB) that concepts are routinely shared not only by different subjects, but also by different time-slices of the same person. (For this reason, I will use the phrase "different subjects" to cover different people as well as different time-slices of the same person).

2.3. Publicity And Holism

Why does publicity matter? Among other things, because it can be used to argue against certain theories of concepts. In particular, the principle has been frequently used by "conceptual atomists" like Fodor against *holistic* versions of Inferential Role Semantics (IRS). Following Fodor⁸³, I will define an IRS (or "inferentialist") theory of concepts as one that holds the following two theses regarding the conditions for concept individuation and possession⁸⁴:

- **Concept individuation (IRS):** a concept is partially individuated by its inferential role: two concepts C_1-C_2 are the same concept only if they have the same inferential role⁸⁵.

⁸³ V. Fodor (1998, pp. 13-14 and 35); Fodor (2004, pp. 32-34). I prefer "Inferential Role Semantics" to the potentially misleading label "Concept Pragmatism", which Fodor has been employing in recent years. Reasons to prefer the former are well summarized by Weiskopf and Bechtel (2004, pp. 48-49) and Glock (2009, p. 6, fn. 2).

⁸⁴ Several authors have defended versions of IRS: v. among others Block (1993, 1998), Harman (1999), Peacocke (1992, 2007), Prinz (2002), Rey (1983, 1985, 2005), Schneider (2011), Weiskopf (2009a).

⁸⁵ Notice that this leaves open whether concepts are *entirely* individuated by their inferential role. For instance, an IRS theorist could hold that concepts are individuated by inferential role *plus* reference, so that sameness of reference is also required for two concepts to be the same concept. I will not discuss the issue here, since what matters for our purposes is the partial individuation claim. V. Schneider (2009 a,b) for discussion.

- **Concept possession (IRS):** having a concept requires having certain specific inferential dispositions: a subject S has a concept C only if he has some mental representation with the same inferential role as C .

For instance: it follows from the individuation claim that my concept DOG_1 and your concept DOG_2 are the same concept only if they have the same inferential role; and it follows from the possession claim that I have the concept DOG only if I have some mental representation C that has the same inferential role as the concept DOG . (As usual, the two claims are closely connected: if concepts are partially individuated by their inferential role, then I can only have concept C if I have some mental representation that has C 's role).

It is of course crucial for any IRS theory to give some account of what the inferential role of a concept is. But different versions of IRS will spell out the notion in very different ways. In particular, holistic versions of IRS individuate inferential roles very finely, while non-holistic ones individuate them more coarsely. So we can say that all IRS views agree on the two claims above, but diverge on how to construe inferential roles.

A holistic IRS theory (from now on, simply “holism”) can be defined as the conjunction of the following theses⁸⁶:

- **Concept individuation (IRS) plus concept possession (IRS) plus:**
- **Holistic definition of inferential roles:** the inferential role of a concept C is the set of *all* the inferential connections in which C stands at a time t for a subject S ⁸⁷.

⁸⁶ Other definitions of “IRS” and “holism” are no doubt available, but this is the version of the view that I am interested in here.

⁸⁷ Cf. the formulation of holism given by Susan Schneider:

[...] symbols must be individuated by their computational roles [...] (Schneider 2009a, p. 545).

[The Computational Theory of the Mind] requires a theory that types tokens by sameness and difference of total computational role, where the total computational role of a symbol is understood as the role it plays in the algorithms of a completed cognitive science (*ibid.*, p. 524).

(It must be noted that Schneider's formulation presupposes a non-Fregean ontology for concepts, whereas I intend my formulation to be compatible with Fregean views as well. V. also Schneider (2009a, pp. 534-36; 540; 548) and Schneider (2011)). Cf. also the statement of holism given by Ned Block:

[According to holism] all the inferences in which an expression participates are included in its inferential role (Block 1993, p. 39).

(Block also defends a parallel holistic claim at the level of mental representations).

What is it for a concept to stand in an inferential connection for a subject at a time? Suppose I have a certain concept (call it DOG_1) and I am disposed to make certain inferences involving that concept at time t . For instance suppose that, since I believe that all dogs are animals, I am disposed to infer from $X \text{ IS A DOG}_1$ to $X \text{ IS AN ANIMAL}$ at t ⁸⁸. We can then say that the concept DOG_1 stands in an inferential connection to the concept ANIMAL for me at t . Assuming a holistic definition of inferential roles, it will then follow that the inference $X \text{ IS A DOG}_1 \rightarrow X \text{ IS AN ANIMAL}$ is part of the inferential role of my concept DOG_1 . Combining this consequence with IRS, the result is that my concept DOG_1 is partially individuated by the inference $X \text{ IS A DOG}_1 \rightarrow X \text{ IS AN ANIMAL}$, and that someone can have DOG_1 only if he is disposed to infer from $X \text{ IS A DOG}_1$ to $X \text{ IS AN ANIMAL}$.

Holistic IRS views like the one just described have been defended by various authors, most notably Ned Block and Susan Schneider⁸⁹. We are now in a position to see how the publicity principle can be used against such views. As Fodor and others have argued⁹⁰, holism entails that two subjects and two time-slices of the same subject will *almost never have any concepts in common*. Suppose for instance that you and I have (to put it a bit roughly) very similar beliefs about dogs: we both believe they are animals, mammals and so on. We will then respectively have two concepts DOG_1 and DOG_2 with very similar inferential roles. For instance, I will be disposed to infer from $X \text{ IS A DOG}_1$ to $X \text{ IS AN ANIMAL}$, just like you are disposed to infer from $X \text{ IS A DOG}_2$ to $X \text{ IS AN ANIMAL}$, and so on for all the other properties we ascribe to dogs. But suppose that, as it might well happen, I am disposed to infer from $X \text{ IS A DOG}_1$ to $X \text{ BELONGS TO THE SAME SPECIES AS ANDREA ONOFRI'S PET}$, while you are *not* disposed to infer from $X \text{ IS A DOG}_2$ to $X \text{ BELONGS TO THE SAME SPECIES AS ANDREA ONOFRI'S PET}$ (maybe you just don't know I have a dog as a pet). If holism is true, then my DOG_1 and your DOG_2 are *different* concepts, since they have different inferential roles at t . Moreover, this will also determine a difference in our respective concepts $\text{ANIMAL}_1/\text{ANIMAL}_2$, $\text{MAMMAL}_1/\text{MAMMAL}_2$ and so on: these concepts are inferentially related to our respective concepts $\text{DOG}_1/\text{DOG}_2$, so they will also have different inferential roles if DOG_1 and DOG_2 are different concepts. Under holism, any slight difference in our beliefs will thus determine a difference in the concepts involved in those beliefs, as well as all the concepts related to them. This means that, in order for one of my concepts to be the same as one of your concepts, we must have the same global set of beliefs about the world. Clearly, this will almost never happen; nor will different time-slices of the same subject ever have the same beliefs, since people change their mind across time. Holism is therefore clearly incompatible with a principle like (PUB), since it entails that two

⁸⁸ Recall that expressions in small capitals refer to concepts: the expression “ DOG_1 ” is thus meant to pick out a certain specific concept I possess and use in my inferences and judgments.

⁸⁹ V. Block (1993, 1995, 1998) and Schneider (2005, 2009 a,b, 2011); v. also Rupert (2008). Later on, we will examine some differences between Block and Schneider's views.

⁹⁰ V. Fodor (1998, p. 29); Fodor and Lepore (1992, pp. 13-16); Laurence and Margolis (2002, pp. 48-49); Laurence and Margolis (1999, pp. 48-50); Aydede (2000a, pp. 16-21).

subjects (and even two time-slices of the same subject) will almost never have any concepts in common. But (PUB) follows from independently plausible RTM principles that the holist himself accepts; so, the Fodoran would conclude, holism must simply be rejected⁹¹.

Before we move on, it will be useful to say a bit more about the notion of “concept individuation” in relation to the two ontologies of concepts that were discussed earlier. In this section, I have presented theories of concept individuation like IRS and holism; but what does it mean to say that a concept C is “individuated” in a certain way, e.g. by all its inferential connections? Again, claims about concept individuation will have to be interpreted differently depending on the ontology of concepts we adopt:

- On a *Fregean* ontology, a theory of concept individuation will spell out the *identity conditions* for concepts, where these are identified with abstract objects expressed by our mental representations. For instance: claiming that concept C is individuated by property F amounts to claiming that a concept C_x is identical to C only if C_x also has F . So suppose we follow the holist and claim that my concept DOG_1 is partially individuated by its inferential connection to BELONGS TO THE SAME SPECIES AS ANDREA ONOFRI’S PET (as well as ANIMAL, MAMMAL etc...). On a Fregean ontology, this amounts to the following claim: a mental representation [dog] expresses the abstract concept DOG_1 only if its owner is disposed to infer from [dog] to [*belongs to the same species as Andrea Onofri’s pet*]. If the subject doesn’t have such an inferential disposition, then that mental representation does not express DOG_1 , but a numerically *distinct* concept DOG_2 : this is because, again, DOG_1 is partially individuated by its connection to BELONGS TO THE SAME SPECIES AS ANDREA ONOFRI’S PET.
- On a *non-Fregean* ontology, a theory of concept individuation will spell out the *typing conditions* for concepts, where these are identified with concrete mental representations. For instance: claiming that concept C is individuated by property F amounts to claiming that two token concepts C_1-C_2 will both belong to type concept C only if they both have F . We will then interpret a holistic theory of concept individuation in the following terms: your token concept DOG_2 belongs to the same type concept DOG as my token concept DOG_1 only if those two token concepts stand in the very same inferential connections. For instance, since I am disposed to infer

⁹¹ Fodor’s anti-holistic argument also plays an essential role in his more general attack on IRS: if holism is incompatible with publicity, then IRS must be rejected altogether, for non-holistic IRS views are affected by equally fatal problems. This makes it even more important to carefully assess Fodor’s publicity argument against holism (although Fodor also has further anti-holistic objections; v. especially the compositionality arguments in Fodor and Lepore 2002, chs. 1-2). For a good reconstruction of Fodor’s “master argument” against IRS, v. Prinz and Clark (2004, pp. 59-60); v. also Fodor (2004, pp. 34-39) and Fodor (2008, ch. 2).

from DOG_1 to $\text{BELONGS TO THE SAME SPECIES AS ANDREA ONOFRI'S PET}$, you must also have that inferential disposition in order for our two concepts to belong to the same type.

This brief overview shows how a theory of concept individuation can always be cast in terms that would be acceptable on both of our ontologies. For this reason, I will often discuss theories of concept individuation in very general terms, without declining them in their Fregean/non-Fregean versions unless it's necessary. This will considerably simplify our discussion: for more details about how to construe Fregean and non-Fregean versions of a theory of concept individuation, I ask the reader to wait until chapter 3 (sect. 4.1), chapter 4 (sect. 3) and chapter 5, where the topic will be extensively discussed.

3. Two Failed Holistic Responses

3.1. Fundamental Generalizations?

In this section I discuss and reject two strategies that have been proposed to defend holism from publicity-based arguments. The first one is suggested by Block (1998). Block grants that, if concepts are holistically individuated, then no two subjects will ever be in the same mental state, but he denies that this should lead us to reject holism:

Fodor and LePore (1992) object to holistic accounts of mental content on the ground that they would preclude psychological laws, for example: the belief that one is in immediate danger causes release of adrenalin. According to holism, there is no such thing as “the” belief that one is in immediate danger because the belief that you designate in this way is not quite the same as the belief that I designate in this way. Beliefs are too fine grained to be referred to in this way. One strategy for dealing with this issue is to observe that many candidate psychological laws can generalize about contents without actually specifying them. Consider this candidate for a law: For *any* action a and any goal g , if one wants g and also believes that a is required for g , then one will try to do a . This is a universally quantified law (because of the role of “any”), albeit a trivial one. Universally quantified laws are a good scientific bet, and these can involve holistic content. By quantifying over goals, one can state laws without committing oneself to two agents ever having exactly the same goal. The point just made says that the holist can allow one kind of psychological law (the quantified kind) but not another (the kind that mentions specific contents such as the belief that one is in danger). But the holist may go further, arguing that there is something wrong with the putative laws of specific contents. The point is that “The belief that one is in immediate danger causes release of adrenalin” stands to psychological law as “Large slippery rocks on mountain-tops can damage cars on roads below”, stands to physical law. Laws should quantify over such specific items, not mention them explicitly.

Let's try to develop the suggestion. Consider a standard psychological generalization:

G2) If a subject S wants to get water and believes that if she opens the fridge she will get water, then other things being equal S will open the fridge.

(G2) simply seems to be a specific instance of the more general:

G) If a subject S wants P and believes that if she performs action a then P, then other things being equal S will perform action a ⁹².

We can see this by noticing that, no matter how we replace the contents of the attitudes to which (G2) makes reference, we still get a true generalization:

⁹² I prefer (G) to the generalization suggested by Block since it conforms to other examples used in this chapter.

G3) If a subject S wants to get [orange juice] and believes that if she opens the fridge she will get [orange juice], then other things being equal S will open the fridge.

We can also change the contents more radically:

G4) If a subject S wants to [annoy x] and believes that [if she opens the fridge she will annoy x], then other things being equal S will open the fridge.

Generalizations (G2)-(G4) seem equally true, predictive and explanatory. Presumably, this is because they are mere consequences of a more fundamental law of intentional psychology, namely (G). So (G) is at the very least a perfectly acceptable intentional generalization. This seems to spell trouble for the publicity principle. (G) is an interpersonally applicable generalization *par excellence*: indeed, it is so general that it seems to apply to *all* the subjects who are covered by the laws of intentional psychology. And yet, (G) doesn't require the subjects it covers to share any *specific* propositional attitude. Two subjects S_1 - S_2 will *both* satisfy (G)'s antecedent, even if one of them only has a desire to get water and the other only has a desire to get juice. So two subjects could both fall under the generalization even if they didn't have *any* propositional attitudes in common. *A fortiori*, (G) doesn't require the subjects it covers to share any specific *concepts*, since no specific *attitudes* are required in order to fall under the generalization. *Contra* (PUB), then, (G) is an interpersonally applicable generalization which does not "involve" any particular concept C . If so, then we have a perfectly acceptable generalization which doesn't require the subjects it covers to share any concepts at all⁹³.

There are several problems with this response to publicity arguments:

Problem n.1: The defender of publicity might argue that even "fundamental" laws of intentional psychology like (G) require the shareability of *some* concepts. After all, the subjects to which (G) applies must still be able to have beliefs whose content has the form: *if I perform action a then P*. But, arguably, these subjects could not have such beliefs if they didn't have the concept of a conditional and the concept of an action. So the shareability of these concepts will be required for the interpersonal applicability of intentional laws like (G), and this is not a constraint that can be satisfied if we assume holism.

⁹³ Of course, the subjects to which (G) applies must be capable of having a desire that P and a belief that if they perform a then P . That is, they must be capable of having *some* desire and *some* belief that performing a will fulfill that desire. But this simply means that they must be able to form beliefs and desires *in general*, i.e. that they must be creatures capable of having intentional states with propositional content. Clearly, this is compatible with holism, which only denies that these subjects could be in the *same* intentional states.

Problem n.2: A further problem is that Block's response might be simply missing the point. Block argues that psychological laws like (G) do not require the shareability of specific mental states. However, as he explicitly concedes, *other* intentional generalizations (e.g. (G2)) do impose that requirement: like (G), (G2) also applies to a large number of subjects, and it could not do so if the *specific attitudes* mentioned in the antecedent (e.g. the desire to get water) were not shareable by all of these subjects. Given RTM, this entails that the subjects in question have certain *specific concepts* (e.g. WATER) in common. Since holism precludes this possibility, holism is incompatible with the interpersonal applicability of laws like (G2) (even though it might be compatible with the interpersonal applicability of more general laws like (G)).

Block would presumably reply that "specific" generalizations like (G2) *cannot* be accepted as laws of intentional psychology: "[...] there is something wrong with the putative laws of specific contents. The point is that 'The belief that one is in immediate danger causes release of adrenalin' stands to psychological law as 'Large slippery rocks on mountain-tops can damage cars on roads below', stands to physical law. Laws should quantify over such specific items, not mention them explicitly" (Block 1998). Block's claim could be supported further by pointing out that (G2) and the like simply follow as direct consequences of the more general (G). Being more fundamental, only (G) and other such "unspecific" generalizations qualify as laws of intentional psychology: specific generalizations like (G2) simply follow from fundamental ones, so they are not genuine laws.

I don't think Block's reply can be made to work. First, it rules out many important psychological generalizations such as "If you see the moon as being on the horizon, then you will see it as oversized". If such generalizations are non-nomological, then most of the generalizations psychologists attempt to discover are not laws. After all, cognitive science is not in the business of stating platitudinous folk-psychological laws like (G); it is rather interested in more specific generalizations like the moon illusion. This is the kind of generalization cognitive scientists put forward once they have gathered the relevant empirical evidence. If the only psychological laws are fundamental generalizations like (G), then most of cognitive science is not in the business of discovering the laws of human psychology.

Second, determining whether (G2) does qualify as a law seems simply irrelevant here. All that matters is that (G2) is a true generalization which applies to more than one subject. If Block grants this much, he will also have to grant that some propositional attitudes are shareable, or else no more than one subject would satisfy (G2)'s antecedent. So the publicity argument would work even if (G2) was not a law, but simply a *true* and *interpersonally applicable* generalization. Now, since Block cannot deny that the generalization is true, he will have to deny that it applies to more than one subject, and hold that only fundamental generalizations like (G) are interpersonally applicable: that is, he will have to maintain that no more than one subject can satisfy (G2)'s antecedent. But this is a very implausible claim: even though they are more

specific than (G), (G2) and the like still seem to apply to several agents, and to have antecedents that can be easily satisfied by people like you and me!

Problem n.3: Suppose, however, that we grant Block the implausible claim that (G2) and the like only apply to one subject. Even then, a further problem will arise once we move from intentional *generalizations* to intentional *explanations*. Following the model of explanation that most defenders of publicity seem to presuppose, I will assume that intentional explanations have a *deductive-nomological* (DN) structure⁹⁴. A DN explanation of an event is a sound deductive argument whose conclusion is a statement describing the event in question (the *explanandum*). A DN explanation will always include a *nomological* premise (a true statement expressing a law) and a set of *descriptive* premises stating the initial conditions from which the *explanandum* originated. An “intentional” explanation will then just be a DN explanation that makes reference to intentional states (in our case, propositional attitudes). For example, here is a possible intentional explanation of why Mary opened the fridge based on our generalization (G2):

Explanation (A)

- 1) If a subject S wants to get water and believes that if she opens the fridge she will get water, then other things being equal S will open the fridge.
- 2) Mary wants to get water.
- 3) Mary believes that if she opens the fridge she will get water.
- 4) Other things are equal.
- 5) Mary will open the fridge.

Let an “explanation schema” be a DN explanation like (A), where all singular terms referring to intentional subjects have been replaced with variables ranging over intentional subjects (more precisely, the variables will range over the domain of the universal quantifier employed in the nomological premise). For instance, the following will be the explanation schema subsuming explanation (A):

⁹⁴ Cf. Hempel (1965), Hempel and Oppenheim (1948); v. Woodward (2010) for an overview. The DN model is not widely accepted anymore, and several alternative theories have been proposed. However, nothing in my argument turns on the details of the model, and my point against Block would also go through assuming a different theory of explanation. Moreover, given his emphasis on fundamental laws like (G), Block presumably endorses at least some aspects of the DN model, like the principle that laws are necessary components of successful explanations.

Explanation schema (A)

- 1) If a subject S wants to get water and believes that if she opens the fridge she will get water, then other things being equal S will open the fridge.
- 2) S wants to get water.
- 3) S believes that if she opens the fridge she will get water.
- 4) Other things are equal.
- 5) S will open the fridge.

Against Block, the defender of publicity might argue that schema (A) is *interpersonally applicable* just like its nomological premise (1) (which, notice, is simply generalization (G2)). Suppose Rosie, Janet, Claire etc... also want to get water, believe that if they open the fridge they will get water, and open the fridge as a result. By replacing the variable “S” in the schema with the names “Rosie”, “Janet”, “Claire”, we will get deductive arguments that have the same form as the explanation (A) we used for Mary. Since the premises and conclusions of the resulting arguments will be true of the subjects in question, each argument will be a sound DN explanation. So explanation schema (A) is applicable to all of these subjects, i.e. it is “interpersonally applicable”.

Of course, the explanations obtained after replacing the variable would not be sound unless the subjects in question *all* satisfied the relevant descriptive premises: the arguments obtained from schema (A) will only explain the actions of Mary, Rosie, Janet, Claire... if these subjects *do* want to get water and believe that, if they open their respective fridges, they will get water. But if these subjects all want to get water, then they have a specific propositional attitude in common, and this is incompatible with holism. Therefore, the defender of publicity will conclude, holism is not only incompatible with the interpersonal applicability of generalizations like (G2), but also with the interpersonal applicability of the explanations based on them.

Following his usual strategy, Block might be tempted to respond that an argument like (A) is *not* an interpersonally applicable explanation, since it is based on a “specific” generalization like (G2) and these generalizations only apply to one subject (v. problem n. 2 *supra*). In fact, Block might go as far as denying that (A) is an explanation at all: if (G2) is not a law, but only a true generalization, then (A) does not count as an explanation under the DN model.

Unfortunately for Block, this reply would miss the point: our argument goes through even on the DN explanation Block would favor, i.e. one based on fundamental laws of psychology like (G). Consider the following, alternative explanation of Mary’s action:

Explanation (B)

- 1) If a subject S wants P and believes that if she performs action a then P, then other things being equal S will perform action a [**notice this is just law (G)**]
- 2) Mary wants to get water.
- 3) Mary believes that if she opens the fridge she will get water.
- 4) Other things are equal.
- 5) Mary will open the fridge.

The explanation schema for (B) is:

Explanation schema (B)

- 1) If a subject S wants P and believes that if she performs action a then P, then other things being equal S will perform action a .
- 2) S wants to get water.
- 3) S believes that if she opens the fridge she will get water.
- 4) Other things are equal.
- 5) S will open the fridge.

Like (A), schema (B) is applicable to multiple subjects (Mary, Rosie, Janet, Claire...) and can be used to produce sound explanations of their actions. Again, this seems incompatible with holism. If the DN arguments obtained by replacing “S” with the names of Mary, Rosie etc... are sound, then descriptive premises (2)-(3) are true of all these subjects. Therefore, the subjects in question have certain specific propositional attitudes in common (e.g. the desire to get water), which is inconsistent with holism. This time, however, the response Block used for explanation (A) is unavailable. The holist *cannot* respond: “Schema (B) is not interpersonally applicable because its nomological premise (1) isn’t”. Block *agrees* that fundamental psychological laws like (G) apply to multiple subjects, since they don’t make reference to specific mental states. Therefore, he cannot respond that the nomological premise of schema (B) is not interpersonally applicable. But if schema (B) *is* interpersonally applicable, the corresponding mental states are shareable and holism is false.

The only option for Block would be to appeal to the descriptive premises (2)-(3). Such premises, he might hold, are “too specific”: they make reference to particular propositional attitudes like the desire to get water, whereas an interpersonally applicable DN argument should only employ unspecific premises such as “S wants P”. But this reply seems even more implausible than the one given in response to problem 2. There, Block had to concede that we can’t apply generalizations like (G2) to more than one subject; it now turns out that we *also* can’t use explanations that make reference to specific beliefs/desires in order to explain the behavior of multiple subjects! Clearly, this claim is even more problematic, as it would rule out most of the DN explanations we actually use as non-interpersonally applicable.

Problem n. 4: In fact, the holist’s troubles have a deeper source. Issues with generalizations and explanations arise from Block’s concession that (under holism) no more than one subject can have a specific attitude *pa* (e.g. the desire to get water, or the belief that dogs bark). Suppose Block is right, and consider the following “schematic ascription” (D_s):

(D_s) S believes that dogs bark

(A schematic ascription is simply a propositional attitude ascription (e.g. “Sam believes that dogs bark”) where the singular term in subject position has been replaced with a variable ranging over intentional subjects). According to Block, holism entails that (D_s) can only be true of a specific subject S (say, Gary). Indeed, if Block is right (D_s) will only be true of Gary at a specific time *t*, since people change their inferential dispositions across time. To put it more simply: if holism is true, then the predicate “believes that dogs bark” only applies to a single time-slice of a specific subject. And this is simply incredible. Holism now appears to be in contrast not only with the interpersonal applicability of generalizations and explanations, but with our most fundamental intuitions about the general applicability of intentional predicates. If holism does have this consequence, then it must clearly be rejected; no metaphysical view about the nature of propositional attitudes can be correct if it leads to such revisionary claims about the semantics of natural language.

The interpersonal applicability of intentional generalizations/explanations, on which anti-holistic arguments have relied, is thus only a manifestation of a deeper problem. The real issue is: holism seems to have the unacceptable consequence that attitude predicates can only apply to one subject at a certain specific time. In turn, of course, this brings about further problematic consequences: no generalization ever has an antecedent that is true of more than one subject, and no schematic explanation has descriptive premises that are satisfied by more than one agent. Once we see that the latter problems have a deeper source, however, we also realize that Block’s solution to them is bound to fail: *even if* Block’s appeal to “fundamental” laws like (G) could help

solve the problem of generalizations/explanations, it would still be unable to account for the interpersonal applicability of intentional predicates⁹⁵.

3.2. The Appeal To Similarity

A different strategy that several holists have found tempting consists in appealing to concept *similarity* rather than concept *identity* to account for the interpersonal applicability of intentional generalizations⁹⁶. (PUB) requires the subjects to which a certain generalization applies to have “the same concept”, i.e. that specific concept which is “involved” in the generalization. A holist might reply that a generalization can apply to multiple subjects as long as they have concepts which are *similar enough*: sameness of concepts is not required.

While seemingly intuitive and unproblematic, the appeal to concept similarity raises a number of challenges for the holist. First, as Fodor has famously argued, the possibility of concept *similarity* seems to be parasitic on the possibility of concept *identity*; if so, the holist is still in trouble, as concept identity is (almost) impossible if concepts are holistically individuated. Fodor’s objection goes along the following lines⁹⁷. Consider the subjects covered by generalization (G1):

(G1) If a subject S wants to get water, then other things being equal S will look for water.

Suppose the holist claims that (G1) can apply to all these subjects as long as their respective mental representations I GET WATER⁹⁸, as well as their constituent concepts, are similar enough. Our driving question will be: *what does the similarity of their concepts for water consist in?*

First option: Their similarity consists in the fact that their inferential roles contain many *identical* inferential connections:

⁹⁵ Susan Schneider has developed a more sophisticated response to publicity arguments, one which combines Block’s strategy with her Millian semantics for ascriptions. I will discuss Schneider’s response in ch. 3 (sect. 4.2).

⁹⁶ V. Block (1998) and Harman (1993, pp. 169-79, quoted by Fodor 1998, p. 30). V. also Laurence and Margolis (1999, pp. 49-50) for discussion.

⁹⁷ V. for instance Fodor (1998, pp. 29-35).

⁹⁸ To simplify the exposition, I switch to small capitals as a notation for mental representations. This is not meant to beg the question against the Fregean ontologist; even though I use the same notation for concepts and mental representations, I want to leave open whether concepts are expressed by mental representations or identical to them.

- S_1 has WATER₁ and is disposed to infer: $X \text{ IS WATER}_1 \rightarrow X \text{ IS LIQUID}_1, \text{TRANSPARENT}_1, \text{ODORLESS}_1 \dots$
- S_2 has WATER₂ and is disposed to infer: $X \text{ IS WATER}_2 \rightarrow X \text{ IS LIQUID}_2, \text{TRANSPARENT}_2, \text{ODORLESS}_2^{99} \dots$

And so on for all the other subjects. The intuitive idea is that the inferential roles of the concepts in question “overlap” enough: they contain many of the *same* inferential connections (as well as some different ones), and this is what accounts for their being similar. Unfortunately, as Fodor notes, this option is not available to the holist. Assuming holism, S_1 ’s LIQUID₁ will be a *different concept* from S_2 ’s LIQUID₂; *mutatis mutandis* for TRANSPARENT, ODORLESS etc.... Therefore, the inferential connection $X \text{ IS WATER}_1 \rightarrow X \text{ IS LIQUID}_1$ included in WATER₁’s inferential role will be *different* from the inferential connection $X \text{ IS WATER}_2 \rightarrow X \text{ IS LIQUID}_2$ included in WATER₂’s inferential role, since they are constituted by different concepts. Or, to put the point differently: the holist would now have to give an account of what the identity of these inferential connections consists in, and it seems he has no way to do so. For instance, suppose he responds that the two inferential connections are the same because the concepts WATER₁-WATER₂, LIQUID₁-LIQUID₂ involved in them have *same reference*. If so, then the holist must concede that WATER₁ is the same concept as WATER₂, and LIQUID₁ the same concept as LIQUID₂: this is what makes the inferential connections in which they are respectively involved *the same* connection. But, of course, if WATER₁ is the same concept as WATER₂, and LIQUID₁ the same concept as LIQUID₂, holism is straightforwardly false, since these concepts have different inferential roles.

Second option: Alternatively, the holist could respond that the similarity of water-concepts consists in the fact that their inferential roles contain many *similar* inferential connections:

- S_1 has WATER₁ and is disposed to infer: $X \text{ IS WATER}_1 \rightarrow X \text{ IS LIQUID}_1, \text{TRANSPARENT}_1, \text{ODORLESS}_1 \dots$
- S_2 has WATER₂ and is disposed to infer: $X \text{ IS WATER}_2 \rightarrow X \text{ IS LIQUID}_2, \text{TRANSPARENT}_2, \text{ODORLESS}_2 \dots$

Where, for instance, the inferential connection between $X \text{ IS WATER}_1$ and $X \text{ IS LIQUID}_1$ is *sufficiently similar* to the inferential connection between $X \text{ IS WATER}_2$ and $X \text{ IS LIQUID}_2$. The problem with this strategy, of course, is that it simply seems to launch a regress. Our original question will now arise again, but at the level of inferential connections: what does the similarity of these two connections consist in? Clearly, the holist cannot appeal to the similarity of the concepts involved to answer this question, or he would

⁹⁹ The indexes indicate the subject to which the concepts in question belong.

now have to give an account of concept similarity that didn't appeal to the similarity of their connections.

Third option: The most promising strategy, it seems, would be to simply refuse to answer our original question and take concept similarity to be *primitive*. The holist cannot be required to give an account of what it is for water-concepts to be similar, since it's impossible to give such an account: roughly speaking, there is *nothing* their similarity consists in, i.e. no other fact to which their being similar can be "reduced". Fodor's response to this move is: "Sure; but then why not take content *identity* as primitive and stop trying to construe *it*? In which case, what is semantics *for*?" (Fodor 1998, p. 32 fn. 5). It is not clear to me what Fodor has in mind here. In any case, there are deeper problems with this response, for the appeal to primitive similarity appears entirely *ad hoc*. *Why* should we take the similarity of two concepts to be a "fundamental fact of the universe" analogous to (say) an electron's having a certain spin?

There seem to be two possible ways for the holist to justify his response. First, he might adopt an *anti-reductionist* position about mental representations in general¹⁰⁰: just like all other facts involving mental representations, the fact that two concepts are similar cannot be fully reduced in non-intentional terms. But not only would this reply be *ad hoc*; it would also miss the point, since the holist is *allowed* to employ intentional notions in giving an account of concept similarity. The holist is not being asked to provide a reductive account of concept similarity in non-intentional terms. He is only being asked to provide some account of what concept similarity amounts to, and he can freely help himself to other intentional notions (such as concept identity) if he wants to. The problem is not that the holist cannot offer a *reductive* account of concept similarity, but that he cannot offer any non-circular account *whatever*.

A more promising strategy would be to adopt a metaphysical position on which similarity facts were *in general* irreducible to identity facts. Trope theorists, for instance, have argued that the similarity of two tropes t_1-t_2 cannot be reduced to their having some property in common, i.e. to their having the *same* property F: their being similar is a fundamental, primitive fact that cannot be further reduced¹⁰¹. Similarly, one might argue that facts about concept similarity cannot be reduced to facts about concept identity. It might then be permissible to refuse to answer our original question, just like the trope theorist will refuse to give an account of the fact that t_1-t_2 are similar. Regardless of whether this would block Fodor's objection, however, appealing to a controversial metaphysical view in order to save concept holism appears dialectically unacceptable in this context. The holist should offer independent reasons for thinking trope theory was true before employing it against Fodor's charge¹⁰².

¹⁰⁰ V. e.g. Burge (2010).

¹⁰¹ V. e.g. Williams (1953).

¹⁰² Indeed, Block (1998) does not take this line and concedes that the holist must offer some account of what concept similarity amounts to.

There is also a second, more general problem affecting every appeal to concept similarity. In a number of ordinary cases, we have no problems saying that two people have “the same *F*”, for some category *F* to which the object in question belongs. For instance: we routinely claim that different people have the same car, the same sweater, the same dressing style and so on. We often use these claims to ascribe an identity in *type* or *kind* to the objects in question: this is what I do if I know that Tim and Susan each have a Rolls-Royce Silver Ghost and claim “Tim and Susan have the same car”. Sometimes, however, we also use sameness claims to ascribe *numerical* identity; this is what I do if I know that Tim and Susan own one and the same car (suppose they are a couple), in which case each subject’s car is numerically identical to the other subject’s car. For most classes of objects, both kinds of sameness ascriptions are entirely unproblematic; there is no reason to think that they will be systematically false due to the peculiar metaphysical nature of cars, sweaters or dressing styles¹⁰³.

Clearly, then, the holist who appeals to concept similarity must think that concepts are *different* from most everyday objects. That is, he must think that concepts constitute a peculiar ontological kind, such that:

- Different subjects cannot have *numerically identical* concepts, and any ascription of numerically identical concepts to distinct subjects is false.
- Different subjects cannot even have concepts of the same *type* or *kind*; again, any ascription of type- or kind-identical concepts to distinct subjects is false.

If the holist’s appeal to concept similarity was correct, all our claims about different subjects having “the same concept” would be strictly speaking false; we would only be entitled to ascribe similar concepts. But, clearly, this position will only be plausible if supported by a suitable metaphysics of concepts, one showing why concepts should constitute such a peculiar class of entities. For if concepts are like other kinds of objects, it should always be possible to ascribe *at least* type-identity to them, and perhaps even strict numerical identity. Of course, this does not mean that two concepts cannot be similar: just like other kinds of objects, they will be (more or less) similar when they have a certain number of properties in common (reference, inferential connections etc...). But then, as with other objects, their similarity will consist in their having many of the *same* properties¹⁰⁴. And if two concepts have the same property *F*, then there will be at least one type to which they both belong, i.e. the type including all and only the *F*s: that is, they will be type-identical¹⁰⁵.

¹⁰³ One might object that two subjects cannot have numerically identical dressing styles. This will depend on one’s metaphysics of styles, and in any case it will still be possible for different subjects to have type-identical styles.

¹⁰⁴ Unless the holist adopts a trope theory for *all* classes of objects, an option I have already rejected.

¹⁰⁵ There are some complications having to with the type-token taxonomy here; v. ch. 3 (sect. 4.1) for discussion.

I conclude that, *unless* the holist has some specific metaphysical reasons to think that *both* numerical identity and type-identity are impossible between concepts, his claim that ordinary subjects only have *similar* concepts will be clearly unacceptable. At the very least, if two subjects have similar concepts there will be one type to which the two concepts both belong, in which case we can truly claim that they have “the same concept” (in the type-identity sense). Together with Fodor’s objection, this shows that the holist’s appeal to concept similarity is not a successful reply to the argument from publicity¹⁰⁶.

¹⁰⁶ Later on (ch. 3, sect. 4.1), I will discuss a possible construal of the appeal to similarity which I find more convincing. However, a lot more background is needed before introducing this version of the strategy, so I ask the reader to wait until then.

4. Back To Publicity

In this section I will consider a different strategy that the holist might employ to deal with publicity-based objections. The strategy consists in rejecting the argument from psychological generalizations as invalid. This possibility is often obscured by standard formulations of the argument, which never specify its assumptions in a sufficiently precise way (as the reader can confirm by looking back at the quotes in sect. 2.2). Partly for this reason, the issue doesn't seem to have been acknowledged in the literature. This is unfortunate, not just because a possible defense of holism has been overlooked as a result, but especially because the issue is important in and of itself. As we will see later, the argument from psychological generalizations seems to rest on a highly controversial interpretation of RTM, and it is all but clear whether we should accept that version of the view. This will leave our main question unanswered: does the interpersonal applicability of psychological generalizations require intentional subjects to have the same concepts? In the next chapter, I will try to provide a new answer to this question by taking a slightly different perspective on the issue.

Let us briefly restate the argument for publicity considered in section 2.2 in a more general form:

- a) An intentional generalization G “applies” to several subjects: that is, there are many intentional agents who satisfy its antecedent and consequent (when other things are equal).
- b) The antecedent and/or consequent of an intentional generalization G embeds a schematic attitude ascription: this is what makes the generalization intentional. (For instance, (G1) embeds the schematic ascription “S wants to get water”). So the agents to which G applies must satisfy the ascription in question: if “*A*” is the relevant propositional attitude verb, then “S *A*s that P” will be true of all the subjects to which G applies. (For instance, “S wants to get water” is true of all the subjects covered by (G1)).
- c) Since the ascription “S *A*s that P” is true of all the subjects to which G applies, these subjects must all have the corresponding propositional attitude *pa*; if they didn't, the ascription would not be true of them. For instance, since “S wants to get water” is true of all the subjects to which (G1) applies, all these subjects must have the desire to get water.
- d) But, according to RTM (I), someone has a propositional attitude *pa* with content P just in case he stands in an appropriate functional relation to *a mental representation r* expressing P. *Therefore*, all the subjects who have the *same*

propositional attitude pa must also stand in the appropriate functional relation to the *same mental representation r*¹⁰⁷. For instance: all the agents who share the desire to get water must also stand in the desire-relation to the same mental representation I GET WATER.

- e) Finally, it follows from RTM (II) that our mental representation *r* is constituted by a set of concepts [C₁, C₂... C_n]. For instance: the concept WATER is one of the constituents of the mental representation I GET WATER. But then all the agents who stand in a functional relation to *r* must have [C₁, C₂... C_n], since the mental representation in question is constituted by these concepts.
- f) Conclusion: all the agents to which G applies have [C₁, C₂... C_n]. So, for every intentional generalization G, there will be a concept C “involved” in G, i.e. some concept C such that all the subjects to which G applies have C; which is just what (PUB) amounts to.

Having summarized the standard publicity argument in a more general form, we can see more easily what might be wrong with it. The problem, a holist could argue, is that (d) does *not* seem to follow from (c) *plus* RTM (I). A holist will grant that, by RTM (I), someone has a propositional attitude *pa* with content P just in case he stands in an appropriate functional relation to a mental representation expressing P. He could deny, however, that all the subjects who have *pa* must *also* stand in an appropriate functional relation to the *same* mental representation *r* expressing P. For suppose there are (at least) two *different* mental representations *r** and *r*** expressing P: I could then have *pa* in virtue of being related to *r**, while you have *pa* in virtue of being related to *r***. We would then have the same propositional attitude by being related to different mental representations. Going back to our example, the subjects to which (G1) applies might have the desire to get water in virtue of standing in the desire-relation to different mental representations expressing the same proposition *I get water*. I might have that desire by being desire-related to [I GET WATER]*, while you have it by being related to a different mental representation [I GET WATER]**, where both representations express the same proposition *I get water*¹⁰⁸.

In sum, a holist could hold the following: mental representations and their constituent concepts are not shareable, but this is compatible with *propositional attitudes* being shared. The same attitude can be had in virtue of having different mental representations and different concepts. This would block all the objections discussed in previous sections: the claim that concepts are holistically individuated would be

¹⁰⁷ As you might have guessed, the italicized part will be important later.

¹⁰⁸ Arguably, the proposition expressed is not the same in this case, since our respective indexical concepts will pick out different individuals. To avoid this complication, the reader should feel free to use an example involving non-indexical concepts (e.g. one involving the mental representation DOGS BARK). I use the water example to follow Fodor (1998).

compatible with the interpersonal applicability of intentional generalizations, explanations and ascriptions. For instance, holism would be compatible with the fact that multiple subjects have the desire to get water, and with the fact that the schematic ascription “S wants to get water” is true of more than one agent. Of course, the holist would still have to deny that different thinkers can have the same concept WATER; that is, he would still have to reject the publicity of concepts. But why would that be a problem, if sameness of concepts is not required in order for an intentional generalization to apply to multiple subjects^{109,110}?

¹⁰⁹ Notice that analogous worries arise for the other publicity arguments sketched in section 2.2:

- One could grant that different subjects can have the same desires, beliefs and so on, while denying that this requires them to have the same mental representations; they might simply have the same attitude by standing in the right functional relation to different representations.
- Similarly, one could hold that successful communication does not require different subjects to entertain the same mental representation, but only the same proposition.
- Finally, one could hold that genuine agreement does require having the same belief, but deny that this also requires accepting the same mental representation.

Now, the next chapter will largely be an attempt to provide a different version of the psychological generalizations argument for publicity. As the reader will note, it's not clear whether the same strategy could also be used to “patch” the arguments above. While I do think there are interesting analogies between the two kinds of arguments that could be exploited for this purpose, I will not discuss them here (although I plan to do so in future work). We will however go back to the other publicity arguments in chapter 4 (sect. 1), where they will play an important role in our discussion.

¹¹⁰ There is a complication with the holistic strategy. Suppose you endorse a Fregean ontology of concepts *and* a holistic theory of concept individuation/possession. Then, it seems, the reply I just described would *not* be available. On a Fregean ontology, concepts are the constituents of the *propositions* expressed by our mental representations. Now, by RTM (I) two subjects can only have a propositional attitude *pa* with propositional content P if they have a mental representation expressing P; but then, assuming a Fregean ontology, they will also have the *concepts* [C₁, C₂, ... C_n] constituting P, since these are the constituents of the proposition in question. For instance: two subjects will not stand in the desire-relation to the same proposition *I get water* unless they both have the concept WATER, which is among the constituents of that proposition.

Now, I do not know of any holists who endorse a Fregean ontology of concepts, but I think it would be easy to rephrase the strategy suggested in the main text in terms that would be acceptable for them. To do so, our Fregean holist would just have to replace RTM (I) with the weaker RTM (I)*:

RTM (I)*: A subject S has a propositional attitude *pa* with **Russellian propositional content P** just in case S stands in an appropriate functional relation to an internal mental representation *r* expressing P.

(Where, recall, a Russellian proposition is an ordered pair <*x*, is F> consisting of an object *x* and a property F). Assuming RTM (I)* instead of RTM (I), it's open to our Fregean holist to hold that different subjects can have the same propositional attitude without having the same concepts. Under RTM (I)*, two subjects can have the *same* attitude as long as they are related to the *same* Russellian proposition, even if they have *different* concepts (e.g. different Fregean senses) for each constituent of that proposition. To simplify the discussion, I will simply assume that holists with Fregean sympathies will endorse RTM (I)* instead of RTM (I) as long as they want to employ the strategy suggested in the main text.

What, then, is needed to make the argument for publicity valid and effective against the holist? Here are two principles that the defender of publicity might add as premises to his argument in order to make it valid (the list is not meant to be exhaustive, but these seem to be the most obvious candidates):

Additional Premise n. 1: Different mental representations never express the same proposition.

This premise entails that, for any proposition P , there is at most one mental representation r expressing P . So, whenever two subjects have the same attitude pa with propositional content P , they must both stand in an appropriate functional relation to r . But then, since concepts are the constituents of mental representations, these subjects will also have to share the concepts which constitute r .

Alternatively, we could establish publicity by appealing to:

Additional Premise n. 2: For every propositional attitude pa with propositional content P , there is a *unique mental representation r* such that r expresses P and a subject S has pa iff S stands in an appropriate functional relation to r .

This premise is, in fact, a stronger version of RTM (I). It holds that, in order to have a certain attitude pa with content P , one must be related to a certain specific representation r : it's not enough to be related to some representation or other expressing P . Again, it will then follow from RTM (II) that, whenever two subjects share pa , they must also share the concepts which constitute that unique representation r which is required in order to have pa ¹¹¹.

Here I will not try to determine whether the above premises are true or not. On the face of it, justifying their acceptance seems no easier than providing an argument for publicity itself. Let me briefly mention a few difficulties that might arise:

- To justify n.1, the defender of publicity would have to commit to a specific view concerning the semantics of mental representations, one on which different representations never express the same proposition. (Interestingly, the view in question would not be an option for Fodor. Fodor holds that the semantic content of mental representations reduces to their referential content, so that CICERO IS BALD/TULLY IS BALD are *different* mental representations expressing the *same* proposition¹¹²).

¹¹¹ Premise n. 2 would follow from premise n. 1 + RTM (I). However, an RTM theorist might want to endorse n. 2 not because he also endorses n. 1, but for independent reasons.

¹¹² V. Fodor (1998, p. 39). V. chs. 4-5 for discussion of Fodor's view.

- It's even less clear what kind of argument could be provided for n. 2. Moreover, endorsing n. 2 would force one to specify, in a principled way, *what* particular mental representation is required in order to have a given attitude. For instance: among all the mental representations expressing the proposition *dogs bark*, which one is required in order to have attitudes about that proposition? What special features does it have, which make it distinct from other representations with that content?

Clearly, settling the status of these two premises would be no easier than settling the status of the publicity principle itself. If so, the holist has a clear dialectical advantage: his opponent will have to do a lot more work before having a sound anti-holistic argument based on publicity. More importantly, however, we seem to have reached a stalemate regarding our original question: *does the interpersonal applicability of intentional generalizations really require concepts to be public?*

In the next chapter, I will consider an alternative way to answer this question. Instead of trying to patch the old publicity argument, I will offer a new version of it, one that does not rely on any of the controversial premises above. My defense of publicity will be based on the *non-Millian, contextualist* semantics for ascriptions which I began to defend in chapter 1. As we will see, that semantics *does* require the subjects covered by an intentional generalization to have certain specific concepts in common. Our main task will then be to spell out more precisely what kind of publicity principle is established by our semantics, and what consequences this will have *vis-à-vis* holistic theories of concepts.

Chapter 3

Publicity, Contextualist-Style

1. Introduction

This chapter will argue for two main theses. The first one is that our best semantics for attitude reports entails a version of the publicity principle: assuming that semantics, if a generalization G applies to a group of subjects then there is a specific concept C such that the subjects covered by G must all have C . I will start (sect. 2.1) by developing further the semantics for reports I began to defend in chapter 1. There, I argued that a theory of attitude ascriptions should satisfy two main *desiderata*: it should take into account the context-sensitivity of reports, and it should be non-Millian. I will then describe a family of contextualist theories that were specifically designed to meet such *desiderata* (sect. 2.2), moving on to consider their implications in section 3. There, I will show how contextualism entails a version of the publicity principle and spell out more precisely what that principle amounts to.

In section 4 I will turn to my second thesis: our new publicity principle is still incompatible with holism, but we can modify standard holistic views so as to make them consistent with the principle (sect. 4.1). This will require weakening some of the holist's original claims, thus raising the question of whether such "watered down" holism would still be a form of holism worth having. (I will defend a positive answer to this question in the next two chapters (chs. 4-5)). I will then conclude (sect. 4.2) by rejecting an alternative holistic strategy recently proposed by Susan Schneider, which consists in appealing to a Millian semantics for generalizations in order to reject the publicity principle.

2. Intentional Generalizations And The Semantics Of Reports

2.1. Two Morals

An important point that has too often been ignored in the literature on intentional generalizations is that they are sentences in a language¹¹³. As such, they have a semantic content which is compositionally constructed from that of their constituents. Among the constituents of an intentional generalization, we will of course find a (schematic) propositional attitude ascription: this is what makes the generalization *intentional*¹¹⁴. For instance, our generalization (G1) ("If a subject S wants to get water, then other things being equal S will look for water") embeds the schematic ascription "S wants to get water" in its antecedent. Now, if a true generalization "applies" to a group of subjects, those subjects will satisfy its antecedent (and its consequent, if other things are equal). When this happens, the schematic ascription embedded in the antecedent will be *true* of such subjects. For instance: the subjects to which (G1) applies all satisfy (G1)'s antecedent, which means that the schematic ascription "S wants to get water" is true of all of them. So: a generalization applies to a group of subjects only if the ascriptions embedded in its antecedent/consequent are true of the subjects in the group¹¹⁵.

This seemingly obvious fact has an important consequence: the conditions for interpersonal applicability of intentional generalizations depend on the truth-conditions of attitude ascriptions. For instance: if the truth-conditions of a certain schematic ascription "S believes that P" are such that it is not true of more than one subject at a specific time, then a generalization embedding that ascription will only apply to that subject at that time¹¹⁶. So, once we know what's required for an ascription to be true of

¹¹³ In fact, this problem affects the literature on *ceteris paribus* generalizations in general; v. Carroll (2010). Schneider (2005, 2011) is a notable exception; my approach has been partially inspired by her attempt to use a Millian semantics for generalizations in order to defend holism (v. sect. 4.2 *infra*).

¹¹⁴ Some intentional generalizations might involve non-propositional mental states, but they can be ignored here.

¹¹⁵ There is a complication: the generalization "If a subject S believes that she is going to be attacked by a bear, then other things being equal S will be scared" applies to me, even though I have never satisfied the antecedent. Notice, however, that I *would* satisfy the antecedent if I was in the appropriate circumstances (e.g. if I happened to see a bear rushing angrily towards me); similarly, other intentional subjects will also satisfy the antecedent when they happen to be in those circumstances. We could therefore say that the generalization is "applicable" to all these subjects, since the ascription in the antecedent *would* be true of them under certain conditions. Consequently, notice that if a theory (e.g. holism) entails that these subjects would *not* satisfy the antecedent under those conditions (e.g. because only one subject at a time can have the relevant belief), that theory is *ipso facto* false. The gist of our discussion of holism in the previous chapter is therefore untouched by this complication; in what follows, I propose to ignore this problem whenever possible in order to make the discussion smoother.

¹¹⁶ Everything I say here also applies, *mutatis mutandis*, to intentional explanations. For instance: if the truth-conditions of a certain schematic ascription are such that it is not true of more than one subject at a

a subject, we will also know more about what's required for a generalization *embedding* that ascription to apply to multiple subjects; establishing the truth conditions of attitude reports can help us establish the conditions that must be satisfied for a generalization to cover a group of agents. Suppose for instance that, according to our best semantics for reports, a certain schematic ascription (A) cannot be true of someone who doesn't have a certain specific concept C; we can then conclude that a generalization embedding (A) will not apply to a group of subjects unless they all have C. So choosing a theory of attitude ascriptions might enable us to answer the question that was left unanswered at the end of the last chapter: does the interpersonal applicability of intentional generalizations require concepts to be shared? Let us, then, go back to attitude ascriptions and their semantics to see whether this can help us solve the problem of publicity.

Two important morals can be drawn from our discussion in chapter 1. The first one is that, in many contexts, utterances of attitude ascriptions will encode information not only about the Russellian proposition believed by the subject, but also about the “way” in which the subject believes that proposition. The second one, which I will soon develop in more detail, is that the truth conditions of attitude reports vary with context. These two morals will (hopefully) provide the key to solve our problem.

Let's start by focusing on the first moral, which is simply the conclusion of my arguments against Millianism. According to the Millian, a belief report of the form “S believes that P” is true iff S stands in the BEL-relation to the Russellian proposition expressed by “that P” in *some* way w^{117} . If w_1 and w_2 are two possible ways of believing P, it is irrelevant for the truth of the report whether S believes P through w_1 or w_2 . For instance: the report “Lois believes that Clark can fly” is true of Lois, since she stands in the BEL-relation to SUPERMAN CAN FLY and this mental representation expresses the Russellian proposition <Superman, can fly>¹¹⁸. The fact that Lois does *not* stand in the BEL-relation to CLARK CAN FLY does not affect the truth-value of the ascription. If Millianism is true, then (for any report of the form “S believes that P”) the way in which S believes the Russellian proposition expressed by “that P” is irrelevant for the truth of the report itself.

Now, I have argued that a Millian semantics will be unable to explain our intuitions about attitude ascriptions; moreover, it will also be unable to account for the role that intentional generalizations have to play in our everyday explanations/predictions of behavior. The failure of Millianism gives us reason to think that, on our best theory of attitude reports, the *way* in which the subject of the report does/does not believe¹¹⁹ the

time, then an *explanation schema* having that ascription as one of its premises will only apply to that subject at that time (v. ch. 2, sect. 3.1 for more details about explanation schemas).

¹¹⁷ Cf. my summary of Salmon's account in ch. 1 (sect. 2.2).

¹¹⁸ Note that, of course, this proposition is *identical* to the Russellian proposition <Clark, can fly>.

¹¹⁹ I will henceforth focus on belief ascriptions and assume that everything I say about them will also hold for ascriptions of desires, intentions, etc....

relevant Russellian proposition *should* be relevant for the truth of the report itself. Reports do not just encode the information that S believes a certain Russellian proposition in some way or other; they also tell us something about *the specific way* in which S believes that proposition (in at least a significant number of contexts).

Now, there are at least two ways to think of ways of believing. Ways might be identified with *mental representations*, e.g. with LoT sentences to which we stand in the belief-relation and through which we believe the relevant propositions¹²⁰. If the anti-Millian adopts this construal, he will take the truth of “Lois believes that Clark can fly” to depend on what mental representation Lois uses when she believes <Superman, can fly>. Ways might also be identified with *Fregean senses*, e.g. with Fregean propositions constituted by Fregean senses picking out the constituents of the relevant Russellian proposition. If the anti-Millian adopts this construal, he will take the truth of “Lois believes that Clark can fly” to depend on the Fregean senses under which Lois believes <Superman, can fly> (more specifically, on whether she ascribes the property of *being able to fly* to Superman/Clark under the sense *Superman* or under the sense *Clark*).

The differences between these possible construals do not matter much for our purposes. No matter how we decide to think of ways, the first moral from chapter 1 can be summarized as follows:

First moral: In at least some contexts, the truth-value of an utterance *u* of an ascription “S believes that P” will depend on whether S believes the Russellian proposition expressed by “that P” *and* on the *concepts* under which S believes that proposition.

This should be acceptable for both the non-Millian theorists I just described¹²¹. The non-Fregean will identify concepts with mental representations (e.g. with LoT symbols); he will then hold that the truth of “Lois believes that Clark can fly” depends on whether she believes <Superman, can fly> under the concept/mental representation SUPERMAN or under the concept/mental representation CLARK. The Fregean, on the other hand, will identify concepts with Fregean senses; he will then hold that the truth of “Lois believes that Clark can fly” depends on whether she believes <Superman, can fly> under the concept/sense *Superman* or under the concept/sense *Clark*.

On both views, concepts will be relevant for the truth-conditions of attitude reports. Clearly, this is straightforwardly inconsistent with Millianism. Our first moral shows that the satisfaction conditions of a report are stricter than the Millian thinks. It’s not enough for a subject to believe the relevant Russellian proposition under any old concept; he must also believe that proposition under a concept satisfying certain specific

¹²⁰ This was the view defended by Braun. It is also (roughly) the view of non-Millians such as Crimmins and Perry (1989) or Richard (1990) (more on this soon).

¹²¹ Crimmins and Perry (1989, p. 710) make a similar point; v. also Sainsbury (2010).

requirements. Predictably, then, non-Millian views will also yield stricter applicability conditions for intentional generalizations. To see why, consider the report¹²²:

(L) Lois believes that Clark can fly.

(I will refer to this sentence as “(L)” throughout). Suppose that the non-Millian truth-conditions for (L) are as follows (for the time being, we can ignore the context-sensitivity of (L); we will go back to this later):

(L) is true iff Lois stands in the belief-relation to the thought CLARK CAN FLY

These truth-conditions make reference to a specific thought and a specific concept CLARK which is among the constituents of that thought. The non-Millian might then characterize CLARK as follows: “The concept CLARK is that unique concept which:

- Refers to Superman/Clark
- Is inferentially connected to [IS CALLED “CLARK”, IS SHY, WEARS GLASSES...]¹²³”

Clearly, it follows from our non-Millian truth-conditions that (L) is false: Lois does *not* stand in the belief-relation to any thought consisting of the singular concept CLARK and the predicative concept CAN FLY (although she does stand in the belief-relation to SUPERMAN CAN FLY). Unlike Millianism, then, this semantics would be able to explain our intuition that (L) is false.

Crucially, our semantics for (L) will have immediate consequences for the applicability conditions of certain generalizations. Assuming that semantics, the schematic ascription:

(L_s) S believes that Clark can fly

will only be true of subjects who accept the thought CLARK CAN FLY, so only of subjects who *have* the concept CLARK in the first place. Now consider a generalization *embedding* (L_s), such as¹²⁴:

G5) If a subject S believes that Clark can fly and believes that Clark just jumped off a skyscraper, then other things being equal S will not be worried.

¹²² This was sentence (4) in ch. 1.

¹²³ As usual, if a concept C is “inferentially connected” to concepts [F, G, H...], C’s owner must be disposed to infer from X IS C to X IS F, G, H... (or, more simply, to accept the thought C IS F, G, H...). Whenever a concept C is inferentially connected to a group of concepts, I will put these concepts in square brackets.

¹²⁴ I follow the order of intentional generalizations from ch. 2.

Of course, a subject S won't satisfy (G5)'s antecedent unless the schematic ascription (L_s) is true of him. Assuming our non-Millian semantics for $(L)-(L_s)$, then, (G5) will only apply to subjects who accept CLARK CAN FLY, so only to subjects who *have* the concept CLARK. On that semantics, the conditions for interpersonal applicability are *strict*: for a generalization to apply to a group of subjects, these will be required to have specific concepts like CLARK. (As the reader will have noted, this is a version of the publicity principle; we will go back to this at length in sect. 3 *infra*).

Clearly, the conditions for interpersonal applicability would be rather *less* strict on a Millian semantics. Suppose someone *doesn't* have the concept CLARK, but only some other concept which, while also referring to Superman/Clark, is not inferentially connected to [CALLED "CLARK", SHY, WEARS GLASSES...]. (Maybe this person only has a concept SUPERMAN with the standard "Superman-ish"¹²⁵ role [CALLED "SUPERMAN", HAS SUPERPOWERS, WEARS A COSTUME...]; he just never heard of "Clark Kent"). While this subject obviously cannot accept CLARK CAN FLY, he can still satisfy (G5)'s antecedent on a Millian semantics, e.g. if he accepts SUPERMAN CAN FLY. So the conditions for interpersonal applicability are looser, since (G5) will also apply to subjects who don't have the concept CLARK at all.

Notice that this has immediate consequences for holism. If concepts are holistically individuated, then no more than one subject can have the concept CLARK; if two subjects had the concept in question, they would have to share all their inferential dispositions. Assuming a non-Millian semantics, then, holism would entail that (G5) is not interpersonally applicable. On a Millian semantics, on the other hand, (G5) will also apply to subjects who don't have CLARK, as long as they believe the Russellian proposition <Superman, can fly> under some concept or other. So holism would be compatible with interpersonal applicability on a Millian semantics¹²⁶: it's bad news for the holist that a Millian semantics for generalizations is out at this point.

Having spelled out the consequences that our first "anti-Millian" moral will have for the applicability conditions of intentional generalizations, let's turn to our second moral.

Second moral: The truth conditions of attitude reports vary with context¹²⁷.

I take our discussion of Millianism to have established that, *in several contexts*, an utterance u of sentence (L) would be *false*. This, of course, is what our intuitions about (L) seem to show, and what our best semantics for ascriptions predicts. At the same time, however, there are strong reasons to think that (L) will be *true* in certain *other* contexts. This point was anticipated in chapter 1 (sect. 5), where we discussed the context-sensitivity of reports like "Glenda believes that Bob Dylan has a beautiful

¹²⁵ I borrow the expression from Braun (2001a, p. 99).

¹²⁶ Indeed, this is an essential component of Schneider's defense of holism: v. sect. 4.2 *infra* for criticisms.

¹²⁷ Saul 1999a, p. 356.

voice” (Saul 1999a). Similar considerations apply to (L). For instance, suppose you know nothing about Superman/Clark. I decide to tell you that there is a superhero in Metropolis who wears a certain costume, is able to fly and has other amazing superpowers. Being enlightened about Clark’s secret identity, I also tell you that this person goes by the name “Clark Kent”; however, I *don’t* tell you that he also goes by the name “Superman”. As it happens, no one else informs you about the superhero’s two names; moreover, you also remain ignorant about his everyday life as a reporter and his attempts to conceal his secret identity. In fact, you are completely unaware of the existence of a shy reporter called “Clark Kent”. You are simply left with the information I gave you when I first told you about the superhero, and inquire no further into the matter. Now suppose you ask me “Does Lois believe that Clark can fly?” and I answer:

(L) Lois believes that Clark can fly.

My utterance of (L) seems true in this context, even though Lois would reject both “Clark can fly” and (L) itself. After all, you only know about the superhero under the name “Clark”, so uttering (L) seems a perfectly acceptable way to communicate to you what Lois knows about him¹²⁸.

In light of our second moral, the correct semantics for belief reports will be a “*contextualist*” one that can explain how the truth conditions of sentences like (L) vary with context: in particular, that semantics must be able to account for all those contexts in which (L) appears *false*, as well as those few contexts in which it appears *true*. Putting together our two morals, then, the account we are looking for is one on which:

- a) In at least some contexts, the truth-value of an ascription of the form “S believes that P” depends on whether S believes the Russellian proposition expressed by “that P” *and* on the *concepts* under which S believes that proposition.
- b) The truth conditions of attitude reports *vary with context*.

2.2. Crimmins’ Theory Of Ascriptions

Several contextualist theories have been proposed in the attempt to satisfy our two *desiderata*, and it would be impossible to examine all of them here¹²⁹. I have therefore decided to focus on the specific contextualist view developed by Mark Crimmins, since

¹²⁸ Further cases are offered by Crimmins and Perry (1989); v. also Kripke’s (1979) “Paderewski” case. If you are puzzled by some aspects of the case, note that all I need here is that there be *some* attitude reports whose truth-conditions are context-dependent. (The details of the Superman/Clark story make it a bit harder to cook up “contextualist” cases involving (L)).

¹²⁹ V. for instance Richard (1990), Forbes (1990, 1993), Schiffer (1992).

it fits particularly well with the general terms of our discussion¹³⁰. However, my arguments in this chapter do not depend on the details of Crimmins' account and they would also go through assuming another contextualist view in its place.

Here is a (much simplified) summary of Crimmins' view¹³¹. According to Crimmins, the “that”-clause in a belief report “S believes that *a* is F” expresses the same proposition as the unembedded sentence “*a* is F”. Moreover, just as on Millian accounts, “*a* is F” expresses the Russellian proposition $\langle a, \text{ is } F \rangle$ constituted by the object *a* and the property *being F*. However, belief reports do not only specify *what* Russellian proposition is believed by a subject, but also *how* (i.e. via what mental representation) that proposition is believed; in particular, they specify what “*notions*” and “*ideas*” constitute the mental representation through which the subject believes the proposition in question. A notion is a singular concept referring to the individual picked out by the singular term in the “that” clause (CLARK is a notion); an idea is a predicative concept expressing the property picked out by the predicate (CAN FLY is an idea). (In what follows, I will mostly focus on notions).

Notions are specified by *contextually supplied*, “*unarticulated constituents*” of the proposition expressed by a belief report. An unarticulated constituent is a propositional constituent which is not represented by any linguistic unit in the sentence. So an utterance *u* of a report will be true of a subject S iff S believes the relevant Russellian proposition through a mental representation satisfying certain contextually specified conditions, where the conditions in question are specified by the unarticulated constituents of the proposition expressed by *u*. Crimmins calls these contextually supplied conditions “*providing conditions*”. Such conditions are determined by the speaker's intentions in the relevant context and they obviously play a crucial role in fixing the truth conditions of a report, since they determine what mental representation S must accept in order for the report to be true.

Notions may be specified in two different ways. A notion may be *provided*, in which case it will be a constituent of the proposition expressed by the report. Alternatively, a notion may be only *constrained*, in which case it will not be among the propositional constituents: in this case, the proposition expressed is just that there is *some notion* which is involved in S's belief and which satisfies the description D that is contextually supplied by the speaker in *c*¹³².

Here is an example to illustrate the distinction. (From now on, I will simply talk of “concepts” instead of talking about “notions” and “ideas”; recall that both notions and ideas are concepts according to Crimmins). Consider our ascription (L) “Lois believes that Clark can fly” as uttered in an ordinary context *c* in which it appears *false*. On

¹³⁰ V. Crimmins and Perry (1989), Crimmins (1992). Following Saul (1999a), I will simply refer to this position as Crimmins' view.

¹³¹ My summary is based on Saul (1999a). Like Saul, I will only discuss those features of the account which will be relevant for our discussion.

¹³² V. Crimmins and Perry (1989, pp. 701-709).

Crimmins' account, (L) might have one of two sets of truth conditions, which would both account for our intuitions about (L)'s falsehood. If our context *c* is one in which concepts are directly *provided*, the utterance will have the following truth-conditions:

(L) (provided-concepts version): (L) is true in *c* iff there is some mental representation Lois accepts which expresses the Russellian proposition <Superman, can fly> and is partially constituted by CLARK.

Where CLARK might be that concept which refers to Superman/Clark and has the standard Clark-ish role [CALLED “CLARK”, SHY, WEARS GLASSES...]. Clearly, given the above truth-conditions, (L) will be *false* in *c*: Lois does not accept any mental representation which expresses <Superman, can fly> and is partially constituted by the concept CLARK.

Alternatively, context *c* might be one in which concepts are only *constrained*, in which case our utterance will have the following truth-conditions:

(L) (constrained-concepts version): (L) is true in *c* iff there is some mental representation Lois accepts which expresses the Russellian proposition <Superman, can fly> and is partially constituted by a concept satisfying description D.

Where D might be: “Is a concept that refers to Superman/Clark and has the standard Clark-ish role [CALLED “CLARK”, SHY, WEARS GLASSES...]”. Again, it follows from these truth-conditions that (L) is *false* in *c*: Lois does not accept any mental representation which expresses <Superman, can fly> and is partially constituted by a concept satisfying description D.

While (L) has the same truth-value in both contexts, notice how in the second one we refer to the concept that must be involved in Lois' mental representation through a description, while in the first one we refer to it directly. The difference can be seen most clearly if we hypothesize that there is *no* concept referring to Superman/Clark and having the standard Clark-ish role. In our second context, (L) will be *false*, since there is no concept satisfying D. In our first context, however, “CLARK” will become an empty name like “Santa Claus” and it will fail to refer: consequently, our utterance of (L) will be neither true nor false (on at least some theories of empty names)¹³³.

Following Crimmins, we will assume (as seems plausible) that there are *both* constraining and providing contexts¹³⁴: speakers sometimes pick out concepts descriptively and sometimes refer to them directly. As Saul (1999a) notes, in both kinds of contexts it will be the “providing conditions” fixed by speakers' intentions that determine what mental representation S must accept in order for the report to be true. In

¹³³ Crimmins and Perry (1989, pp. 701-706).

¹³⁴ Crimmins and Perry offer reasons to acknowledge both kinds of contexts, a topic I cannot discuss here: v. Crimmins and Perry (1989, pp. 701-706).

constraining contexts, those conditions (e.g. *having the standard Clark-ish role*) will figure in description D and appear among the propositional constituents. In providing contexts, the conditions will not appear in the proposition, but they will still determine what concept (if any) the speaker is referring to: in the above example, the concept CLARK is *that concept which has the standard Clark-ish role*, i.e. which satisfies the conditions provided by the speaker in that context. Following Saul (1999a, p. 361), we can therefore offer a *single* set of truth-conditions for constraining and providing contexts:

Contextualism: An utterance u of a belief report “S believes that P” is true in context c iff there is *some mental representation r* S accepts which expresses the Russellian proposition expressed by “that P” and is partially constituted by the **concepts** specified by u ’s **providing conditions**¹³⁵.

For instance, (L)’s truth-conditions will be:

Contextualism for (L): An utterance u of (L) is true in context c iff there is some mental representation r Lois accepts which expresses <Superman, can fly> and is partially constituted by the concepts specified by u ’s providing conditions.

Crimmins’ contextualist account yields the right predictions about those ordinary contexts in which reports like (L) appear false, as well as those special contexts in which they appear true. In an ordinary context, speakers will supply the following providing conditions: r must be constituted by a concept having the standard Clark-ish role [CALLED “CLARK”, SHY, WEARS GLASSES...]. Since Lois does not accept a representation r expressing <Superman, can fly> and satisfying this condition, (L) will be false in these contexts (be they providing or constraining; *v. supra*).

In a context where (L) appears true, the providing conditions will be different: in particular, it will *not* be a condition on the truth of the report that r be constituted by a concept with the Clark-ish role. For instance: in the context described earlier, I (= the speaker) clearly *don’t* intend to communicate to you that Lois believes <Superman, can fly> under a concept that is inferentially connected to [CALLED “CLARK”, SHY, WEARS GLASSES...] (the reason why I’m using “Clark” in my report is that this is the only name you have for Superman). Therefore, it is *not* a condition on the truth of the report that

¹³⁵ What about a Fregean who endorsed a contextualist view on which a report is true iff S stands in the belief-relation to a *Fregean descriptive proposition* whose features are contextually specified? This view would also be captured by our formula; we would simply have to feed in a description of the relevant Fregean proposition among the providing conditions, and S would then be required to accept a mental representation expressing that proposition. The equivalence of Fregean and non-Fregean versions of contextualism is noted by Crimmins and Perry (1989, p. 710).

Lois believe <Superman, can fly> under that concept. So (L) can still be true of Lois in this context, which is just what we want.

3. A Contextualist Picture Of Publicity

By appealing to contextually shifting conditions on concepts, a non-Millian contextualist semantics like Crimmins' can account for the contextual variability in our intuitions about reports (second moral). Moreover, it can also account for the fact that, in at least some contexts, a report will be true only if the subject believes the relevant Russellian proposition under a certain specific concept (first moral): this will be that concept which satisfies the conditions provided by the speaker in that context.

In the rest of this chapter, I will assume that some version of the contextualist semantics I just sketched is correct¹³⁶. My plan is to apply that semantics to our two guiding questions, which were left unanswered after we examined the shortcomings of the original publicity argument in the previous chapter:

- Does the interpersonal applicability of intentional generalizations require concepts to be public? More precisely: is it the case that, for every generalization G that applies to two or more subjects, there is a concept C that these subjects must all have in order for G to apply to them? And, more generally: what conditions must be satisfied in order for a generalization to cover multiple agents?
- Is holism incompatible with the interpersonal applicability of intentional generalizations?

In light of our semantics, my answer to the first question will be “yes”: contextualism does entail a version of the publicity principle for intentional generalizations. My answer to the second question (which I will defend in sect. 4.1 *infra*) will be: “yes; but there is a way to modify holism so as to make it compatible with the interpersonal applicability of generalizations”.

As anticipated, our choice of a semantics for reports can affect the conditions under which a generalization will apply to a group of subjects: depending on our background theory of ascriptions, those conditions will often change dramatically. A contextualist semantics will give us the following picture of those conditions:

- a) *In many contexts*, an intentional generalization will only apply to a group of subjects if they all have a specific concept C .
- b) The conditions under which an intentional generalization applies to a group of subjects *vary with context*.

¹³⁶ I won't be able to discuss objections against contextualism here, although I plan to do so in future work: v. Saul (1999 a,b), Sider (1995), Soames (1995), Clapp (1995), Schiffer (1992, 1996).

Let me address each of these points in turn. According to our general statement of contextualism, a report “S believes that P” will be true of S in context c only if S believes the relevant Russellian proposition under a concept satisfying the conditions provided by the speaker in c . This has immediate consequences for intentional generalizations, which will *only* apply to subjects who *have* a concept that satisfies the contextually supplied conditions. This is because intentional generalizations embed schematic ascriptions in their antecedents/consequents, and these schematic ascriptions will only be true of subjects who have a concept satisfying the relevant conditions.

For illustration, consider a context c in which (L) is uttered. In c , the schematic ascription (L_s) “S believes that Clark can fly” will only be true of subjects who satisfy the conditions associated with (L). For instance, suppose c is one of the many ordinary contexts in which speakers intend (L) to be true only if Lois believes <Superman, can fly> under the concept CLARK. It follows that, in c , (L_s) is only true of a subject S if S believes <Superman, can fly> under CLARK. Now suppose our intentional generalization (G5) is uttered in c :

G5) If a subject S believes that Clark can fly and believes that Clark just jumped off a skyscraper, then other things being equal S will not be worried.

Since (G5) embeds (L_s) in its antecedent, only subjects who believe <Superman, can fly> under CLARK will satisfy (G5)’s antecedent. *A fortiori*, it’s only subjects who *have* CLARK that can satisfy the antecedent of (G5) as uttered in c ¹³⁷. So an utterance of (G5) in c will *only* apply to subjects who possess the concept CLARK. (*Mutatis mutandis* for an explanation schema based on (G5). An utterance of that schema in c will only apply to subjects who have CLARK: if they didn’t, the descriptive premises of the schema would be false of them)¹³⁸.

¹³⁷ (I.e. the antecedent of (G5)’s utterance in c).

¹³⁸ One might be skeptical about the applicability of a contextualist semantics to intentional generalizations and explanations. While sentences like (L) will indeed be uttered by actual speakers in a number of circumstances, generalizations like (G5) presumably won’t be (setting aside highly sophisticated philosophical contexts). Nor will a speaker ever utter a full DN explanation schema based on (G5)!

These are genuine concerns, but I cannot fully address them here. It’s a notorious problem for the DN model in general that speakers rarely (if ever) provide complete deductive-nomological explanations. But then in what sense does my utterance u “Mary wants to get water” explain why she opened the fridge? The standard answer is that u is an “elliptical” explanation “grounded” in an ideal explanation that, while not actually asserted, could be provided in that context (v. Woodward 2010). A generalization will then play an explanatory role even though it is not actually asserted, and so will the other components of the explanation which are left implicit.

Of course, one might be skeptical about the appeal to elliptical explanations. Since both friends and foes of publicity appeal to the DN theory, however, we are allowed to take the move at face value in this context (v. ch. 2, sect. 3.1, fn. 94). Moreover, alternative models of explanation might also face analogous problems. I will therefore assume that, if a contextualist semantics is true for reports like (L), it will also apply to generalizations like (G5) and to explanation schemas based on them. Such

More generally: given a contextualist semantics, a generalization G uttered in context c will only apply to those subjects who have a concept satisfying the conditions that, in c , are associated with the report(s) embedded in the generalization. As we'll now see, this means that, in many contexts, sharing of specific concepts is required, so that a “contextualist” version of the publicity principle is established.

Let's start with those contexts in which concepts are directly *provided*. In these contexts, speakers refer directly and non-descriptively to specific concepts, which must be involved in the belief of the relevant subject in order for the report to be true. The context c that was discussed in the previous two paragraphs is one of these contexts. In c , our speaker is supplying the following truth-conditions for (L) :

(L) is true in c iff there is some mental representation Lois accepts which expresses the Russellian proposition \langle Superman, can fly \rangle and is partially constituted by CLARK.

Where, again, CLARK is that concept which refers to Superman/Clark and has the standard Clark-ish role [CALLED “CLARK”, SHY, WEARS GLASSES...]. Clearly, c is one of the many ordinary contexts in which (L) is false. Crucially, it is also a context in which certain Clark-related generalizations require the subjects they cover to have a *specific concept* in common: that concept being CLARK, of course. For the reasons given in previous paragraphs, a generalization like $(G5)$ (as uttered in c) will not apply to two subjects S_1 - S_2 unless they *both* have CLARK. If they didn't, they could not possibly believe \langle Superman, can fly \rangle under CLARK; they would then fail to satisfy the conditions imposed by the speaker, (L_s) would be false of them, and $(G5)$ would not apply. More generally: for all those contexts in which specific concepts are provided, generalizations uttered in such contexts will require the subjects they cover to have certain specific concepts in common.

Things are a bit different in those contexts in which concepts are *constrained*: here, the speaker will simply provide a description D which must be satisfied by the subject's concepts. If c was one of these contexts, our report (L) might have the following truth-conditions in c :

(L) is true in c iff there is some mental representation Lois accepts which expresses the Russellian proposition \langle Superman, can fly \rangle and is partially constituted by a **concept** satisfying the **description D**: “*Is a concept that refers to Superman/Clark and has the standard Clark-ish role [CALLED “CLARK”, SHY, WEARS GLASSES...]*”

generalizations/explanations embed the schematic ascription (L_s) : therefore, whatever the truth conditions for an utterance of (L) in c are, the same truth conditions will be given (*mutatis mutandis*) for (L_s) in c , and consequently for all those generalizations/explanations which *embed* (L_s) and *might* be uttered in c .

(Again, (L) will be false in c). Since speakers do not directly provide specific concepts in this kind of context, possession of *some* concept satisfying the description will be enough for purposes of interpersonal applicability. For instance, (L_s) can be true in c of anyone having some concept C satisfying D . So suppose we respectively have two concepts C_1-C_2 which both have the standard Clark-ish role. Suppose, moreover, that I believe \langle Superman, can fly \rangle under C_1 and you believe \langle Superman, can fly \rangle under C_2 . (L_s) will then be true of both of us, and we will consequently both satisfy (G5)'s antecedent, *even if* my concept C_1 is *different* from your concept C_2 . So it *seems* that, in constraining contexts, interpersonal applicability will not require intentional subjects to share a specific concept C : any concept satisfying D will do.

But this conclusion would be too hasty: in many constraining contexts, sharing of specific concepts will still be required. The first thing to note is that, in some sense, intentional subjects will still be required to have “the same concept”, i.e. a concept satisfying description D : they will be required to have a concept belonging to the *same* contextually specified *class* of concepts, i.e. the class including all and only the concepts that satisfy D . So a weaker, “descriptive” version of the publicity principle still holds in constraining contexts. Secondly, one could argue that, in many/all constraining contexts, there will be *only one concept* satisfying D : for instance, that there is a *single* concept CLARK which has the standard Clark-ish role, so that concepts C_1-C_2 in the above example are in fact the same concept CLARK¹³⁹.

While important, these considerations do not conclusively show that constraining contexts require concept publicity. After all, someone (e.g. a holist) with a fine-grained theory of concept individuation could grant that intentional subjects must have some concept belonging to the same (contextually specified) class, but deny that the concepts in the class will therefore be “the same concept”. Similarly, this theorist could deny that description D is only satisfied by a single concept: there are multiple concepts satisfying D , e.g. multiple concepts with the standard Clark-ish role. (A holist, for instance, will hold that these concepts are all distinct since they have different global roles, even though they all share the Clark-ish role).

Setting this controversy aside, a different line of argument shows that, in at least *some* constraining contexts, sharing of *specific* concepts will certainly be required. This is because, in many of these contexts, the supplied description D will make reference to *other concepts* that the believer is required to have. Indeed, the above context c is a case in point. In c , (L_s) is only true of subjects who have some concept with the role specified by description D : [CALLED “CLARK”, SHY, WEARS GLASSES...]. But, clearly, someone will have a concept with that role only if he also has the specific concepts included in D , e.g. SHY and GLASSES: if he didn't have such concepts, he certainly couldn't have a concept which was inferentially connected to them! Consequently, our schematic ascription (L_s) will only be true of subjects who have the specific concepts in

¹³⁹ As usual, this claim would have to be declined differently depending on our background ontology: more on this soon.

question. It follows that, within our concept-constraining context, a generalization embedding (L_s) will not apply to two subjects S_1 - S_2 unless they both have the concepts mentioned in the description: for instance, two subjects will both need to have SHY in order for (G5) to apply to them. Therefore, in many¹⁴⁰ concept-constraining contexts a generalization will require intentional subjects to share certain specific concepts, i.e. those that the relevant description D makes reference to.

Of course, this would not happen if constraining descriptions never made reference to other concepts, but this seems extremely implausible. Presumably, speakers will often pick out the concept they are interested in through its connections with other concepts. Surely, links to action and perception might also be relevant: we might e.g. require Lois to accept a mental representation that usually causes utterances of “Clark can fly” and not “Superman can fly”. Still, it seems we would not be able to impose *all* of our contextual constraints in a purely “behavioristic” way, without referring at all to other conceptual representations.

Summarizing the results of our discussion: in *all* contexts in which concepts are *provided*, and in at least *many* contexts in which they are descriptively *constrained*, a generalization will only apply to a group of subjects if they have a specific concept C in common. We have thus established a “contextualist” version of the publicity principle.

Let’s reconstruct in more detail what that principle amounts to, and how the general argument from contextualism to publicity is supposed to go. Consider an utterance u in context c of a belief report (R) “A believes that P”:

- a) Assuming contextualism, u is true iff A believes the Russellian proposition expressed by “that P” under a concept satisfying u ’s providing conditions.
- b) For *all* providing contexts, and for at least *many* constraining contexts, there is a concept C such that a subject has a concept satisfying u ’s providing conditions only if he has C.

(In providing contexts, C will be that very concept which satisfies u ’s providing conditions; in constraining contexts, it will be one of the concepts included in D).

- c) So, for all providing contexts and for many constraining contexts, there is a concept C such that A has a concept satisfying u ’s providing conditions only if he has C [from (a), (b)].
- d) The schematic ascription $(R_s)^{141}$ “S believes that P” is true of a subject S in c only if S has a concept satisfying u ’s providing conditions in c .

¹⁴⁰ (I leave open the possibility that, in *some* constraining contexts, D does not make reference to specific concepts).

¹⁴¹ Of course, (R_s) is simply the schematic ascription derived from our report (R).

(For instance: (L_s) is true of a subject S in c only if S satisfies the providing conditions that (L) : “Lois believes that Clark can fly” has in c).

- e) So, for all providing contexts and for many constraining contexts, there is a concept C such that (R_s) is true of S in c only if S has C [from (c), (d)].
- f) A generalization G embedding (R_s) applies to a group of subjects in c only if (R_s) is true of these subjects in c .

(For instance: (G5) applies to a group of subjects in c only if (L_s) : “S believes that Clark can fly” is true of these subjects in c).

From (e)-(f), our contextualist version of publicity follows:

Contextualist Publicity: For all providing contexts and for (at least) many constraining contexts, there is a concept C such that, in one of these contexts, certain generalizations apply to a group of subjects only if these subjects all have C.

We have thus established that, according to contextualism, the conditions for interpersonal applicability of a generalization will often involve possession of specific concepts; a second thing to note is that those conditions will *vary with context*. This follows straightforwardly from the context-dependence of attitude reports. Consider those non-ordinary contexts in which (L) is *true* (v. the example in sect. 2.1 *supra*). Clearly, in these contexts the speaker does *not* require Lois to believe *<Superman, can fly>* under CLARK, or (L) would be false (where, as usual, we take CLARK to be that concept which refers to Superman/Clark and has the standard Clark-ish role). Now consider subjects who don’t have the concept CLARK at all (maybe they just don’t know about the shy reporter called “Clark Kent”). In one of our non-ordinary contexts, the schematic ascription (L_s) can be true of these subjects even though they lack CLARK: consequently, (G5) could still apply to them and we would be able to use it to explain/predict their behavior. In an ordinary context where CLARK was provided as part of the truth-conditions, none of this would be possible.

In general, a contextualist framework allows for great flexibility in the applicability conditions of a generalization: those conditions will be stricter in some contexts, looser in others. For instance, there might be contexts in which someone who utters (L) only requires Lois to believe *<Superman, can fly>* under a concept that has *part* of the standard Clark-ish role, e.g. under a concept that is connected to [CALLED “CLARK”]¹⁴². This would again be a context in which (L) is false, but the applicability conditions of (G5) would be looser than in other ordinary contexts: the generalization would now apply to subjects who were acquainted with Superman/Clark under the name “Clark

¹⁴² These are what Crimmins calls “*de dicto* notions”: v. Saul (1999a, p. 361).

Kent” without knowing that he is a shy reporter with glasses. In other contexts, the conditions might be even looser. A speaker might intend (L) to be true just in case Lois believes *<Superman, can fly>* under *some* concept referring to Superman/Clark. In this context, (L)’s truth-conditions would be extensionally equivalent to those provided by Millian accounts, and (L) would be true. (G5) would then apply to *anyone* who had a concept referring to Superman/Clark, no matter how impoverished or mistaken their beliefs about Superman/Clark might be. Conclusion: under contextualism, the applicability conditions of a generalization will vary across a wide range of contexts, depending on how strict the contextually provided conditions on concepts are.

Two additional remarks are in order at this point. First: this picture of intentional generalizations will have to be declined differently depending on our background ontology of concepts. Consider those contexts (providing or constraining) in which possession of a specific concept C is required for a generalization to apply to a group of subjects [S₁, S₂ … S_n]. A *Fregean* contextualist will hold that each of these subjects must have some concept [C₁, C₂ … C_n] that is *numerically identical* to the contextually specified C; by the transitivity of identity, it will then follow that [C₁, C₂ … C_n] are also numerically identical to each other. A *non-Fregean* contextualist, on the other hand, will simply hold that each of these subjects must have some *token concept* [C₁, C₂ … C_n], such that each of these concepts belongs to the same contextually specified *type concept* c. On the first picture, speakers contextually provide conditions on the abstract concept that [S₁, S₂ … S_n] must have (e.g. the abstract concept CLARK, which is that concept of Superman/Clark that has such-and-such inferential role). On the second, they provide conditions on the type concept to which each subject’s token concept must belong (e.g. the type concept CLARK, which is that type concept including all and only the token concepts that refer to Superman/Clark and have such-and-such inferential role).

Second: I can now explain why my arguments in this chapter would still go through (with some modifications) if we assumed a Pragmatic Millian theory of ascriptions instead of a non-Millian one¹⁴³. According to PM, utterances of (L) *pragmatically convey* a proposition concerning the *way* in which Lois believes *<Superman, can fly>*: since the proposition in question is false, speakers mistakenly take (L) itself to express a false proposition. As we have seen, the conveyed proposition can be construed in various ways¹⁴⁴: on all of them, however, that proposition is best seen as encoding information about the *concept* under which Lois allegedly believes *<Superman, can fly>*. On Soames’ proposal, the conveyed proposition is something like the descriptive proposition *Lois Lane believes that Clark, the milquetoast reporter, can fly*. This proposition is true just in case Lois believes *<Superman, can fly>* under the concept THE MILQUETOAST REPORTER. On Salmon’s original proposal, the conveyed proposition was something like *Lois Lane believes the proposition that Superman/Clark can fly under*

¹⁴³ As anticipated in ch. 1 (sect. 5).

¹⁴⁴ V. ch. 1, sect. 2.3.

the guise of the sentence “Clark can fly”. As we have seen, however, identifying ways of believing with linguistic items gives rise to various problems¹⁴⁵. Suppose Lois is only acquainted with Superman/Clark “demonstratively”: she ignores the name of both the mighty superhero she has seen flying and of the reporter with glasses she has seen in the office. Clearly, Lois would accept *neither* “Superman can fly” *nor* “Clark can fly”. Now, an ordinary speaker who knew all this would *still* think that “Lois believes that Superman can fly” is true and “Lois believes that Clark can fly” is false. But if belief reports conveyed propositions about linguistic guises, our speaker should judge *both* reports to be false: Lois does not believe the proposition that Superman/Clark can fly under either “Superman can fly” or “Clark can fly” (by hypothesis, she accepts neither of these sentences). The proposition conveyed by (L) on Salmon’s account is thus best seen as concerning the mental (rather than linguistic) representations under which Lois believes <Superman, can fly>. For instance, utterances of (L) might convey: *Lois Lane believes the proposition that Superman/Clark can fly under the guise of a mental representation which is partially constituted by the concept CLARK*¹⁴⁶.

If attitude ascriptions systematically convey propositions about concepts, they will do the same when embedded within intentional generalizations¹⁴⁷. On Salmon’s account, for instance, utterances of (G5) might convey a proposition whose antecedent has the following first conjunct:

If a subject S believes that Superman/Clark can fly under the guise of a mental representation which is partially constituted by the concept CLARK ...

Crucially, this means that the generalization¹⁴⁸ which is *conveyed* by (G5) is one whose antecedent can *only* be satisfied by subjects who have CLARK. Therefore, there is a specific concept that must be shared by all the subjects who fall under the generalization in question. (On the contrary, the antecedent of the generalization which is *semantically expressed* by (G5) will be satisfied by anyone who believes <Superman, can fly>, even if they don’t have CLARK at all). Pragmatic Millian theories will then also entail a version of the publicity principle, one which (unlike the contextualist version stated above) will concern the proposition conveyed by a generalization rather than the one

¹⁴⁵ V. ch. 1, sect. 5.

¹⁴⁶ As Braun (1998, p. 568) notes, Kripke’s Paderewski case offers further reasons not to identify ways with linguistic representations. PM cannot explain our intuitions about belief reports concerning Peter by appealing to pragmatically conveyed metalinguistic propositions, since there is only one natural language name Peter has for Paderewski, i.e. the name “Paderewski”. On the other hand, Peter arguably has two distinct concepts PADEREWSKI₁-PADEREWSKI₂: if the conveyed propositions concern such concepts, we can explain why speakers accept “Peter believes Paderewski has musical talent” in some contexts (those in which we are discussing Peter’s beliefs about Paderewski-the-pianist) but reject it in others (those in which we are discussing Peter’s beliefs about Paderewski-the-politician).

¹⁴⁷ Indeed, this gives PM a way to account for our “use” of intentional generalizations which is not available to the non-pragmatic Millian: v. ch. 1 (sect. 4) for more details.

¹⁴⁸ Note that I am using “generalization” to refer to a proposition rather than a sentence here.

that is semantically expressed. Simplifying a bit, the principle could be stated thus: for any intentional generalization G , there is a concept C such that *the generalization which is pragmatically conveyed by utterances of G* applies to a group of subjects only if these subjects all have C .

4. Holism In A Contextualist Framework

4.1. A Revised Version Of Holism

In response to the first of our two driving questions (“Does the interpersonal applicability of intentional generalizations require concepts to be public?”), I have argued for the following claim: for many contexts (both providing and constraining), there is a concept C such that, in one of these contexts, the utterance of a generalization G will only apply to a group of subjects if these subjects all have C . I have then spelled out further the implications that contextualism will have for the applicability conditions of intentional generalizations, noting how one and the same generalization will have radically different applicability conditions in different contexts.

In conclusion: even if the standard RTM argument for publicity (ch. 2) is invalid, there is an alternative route to publicity, one based on our best semantics for attitude reports. That semantics forces us to recognize that the interpersonal applicability of generalizations *does* require *specific* concepts to be public, in a significant number of contexts. This leads us to our second question: is holism *compatible* with the interpersonal applicability of intentional generalizations across the various contexts in which they are used? If we stick to the original formulation of holism provided in chapter 2 (sect. 2.3), the answer is a clear “no”. According to the holistic theories defended by authors like Block and Schneider, two subjects will have the same concept C only if they have exactly the same inferential dispositions; for instance, this will be required for two subjects to have the same concept CLARK. Now consider all those contexts in which having CLARK is required for a subject to satisfy the schematic ascription (L_s) , either because CLARK is directly provided or because it is part of the constraining description D . In all these contexts, (L_s) cannot be true of more than one subject: consequently, generalizations embedding (L_s) will not apply to multiple subjects when uttered in these contexts. Assuming contextualism, holism turns out to be incompatible with the interpersonal applicability of intentional generalizations in a significant number of contexts.

When conjoined with a contextualist semantics, the original Block-Schneider formulation of holism has unacceptable consequences. Fortunately, some of the holist’s original theses can be revised to avoid these consequences while preserving the spirit of the view¹⁴⁹. Depending on whether our holist endorses a Fregean or a non-Fregean ontology, the details of the strategy will vary: I’ll discuss each version in turn.

Revising holism (non-Fregean version): Consider a type concept C as a class of token concepts. That class will have certain “membership conditions”: a token concept C_x will

¹⁴⁹ An analogous move is suggested by Laurence and Margolis (1999, p. 76, fn. 30) and (in passing) by Block (1993, pp. 56–57). I am much indebted to both of these passages for some of the ideas in this thesis.

belong to C iff it satisfies those conditions (e.g. iff it has a certain reference, a certain inferential role and so on). Clearly, a token concept can belong to multiple type concepts¹⁵⁰, and two token concepts can belong to the same type concept C while also belonging to two *different* type concepts C_1-C_2 . Similarly, my cat and your dog belong to the same kind *mammal* while also belonging to two different species.

This allows the holist to deal with problematic contexts in the following way. Consider a context c in which a specific concept is directly provided, e.g. one in which our ascription (L) has the following truth-conditions:

(L) is true in c iff there is some mental representation Lois accepts which expresses the Russellian proposition <Superman, can fly> and is partially constituted by CLARK.

In this context, having CLARK is required in order to be covered by generalizations like (G5). To account for this, a holist might take the above statement to refer to a type concept CLARK *whose membership conditions are non-holistic*. More precisely, he might rephrase the above truth conditions as:

(L) is true in c iff there is some *token* mental representation Lois accepts which expresses <Superman, can fly> and is partially constituted by *a token concept belonging to the type concept* CLARK.

Where CLARK's membership conditions are non-holistic. (The membership conditions for a type concept C will be non-holistic just in case two token concepts can belong to C even if they don't have the same global inferential role). A holist could for instance take CLARK's membership conditions to include: *referring to Superman/Clark Kent* and *having the standard Clark-ish inferential role*. (In general, the holist can always point at speaker's intentions to provide the right membership conditions for the relevant non-holistic types. In many ordinary contexts, the conditions provided by the speaker will be the ones I just mentioned, although there will also be contexts involving more coarse-grained types: v. sect. 3 *supra*). Two subjects who possess token concepts referring to Superman/Clark and having the standard Clark-ish role will then be able to satisfy (L_s), and fall under the corresponding generalizations, *even if* those token concepts differ in some of their inferential connections (one subject thinks Clark is handsome, the other one doesn't). In this context, given the speaker's communicative intentions, such differences are intuitively irrelevant for the truth of the report and the applicability of the corresponding generalizations; fortunately, the holist can account for this by holding the type concept provided by the speaker to be non-holistic.

¹⁵⁰ The point is made by both Laurence and Margolis (1999, p. 76, fn. 30) and Block (1993, pp. 56-57). V. *infra* for possible objections.

Of course, this strategy is *consistent* with the claim that some type concepts *do* have holistic membership conditions. For instance, consider the type concept CLARK_H comprising all and only those token concepts that refer to Superman/Clark and have the inferential role that my token concept for Clark has right now. Clearly, no one other than my current time-slice will have a token concept belonging to CLARK_H . Luckily, a token can belong to multiple types at the same time: while our respective token concepts for Clark cannot both belong to CLARK_H , they *do* belong to the same, more coarse-grained type CLARK . Similarly, my cat and your dog both belong to the kind *mammal* while also belonging to two different species; and just like the species *dog* is encompassed by the broader kind *animal*¹⁵¹, CLARK_H will be a more fine-grained type encompassed by the broader type concept CLARK ¹⁵². Since it's the latter, non-holistic type that the speaker refers to in the above context, we can both fall under generalizations uttered in that context. A holist can thus account for interpersonal applicability in providing contexts while maintaining that our token concepts *do* belong to distinct holistic types!

A similar line will be applied to *constraining* contexts in which specific concepts are part of the contextually supplied description D; v. for instance the concept *SHY*, which will be part of D in most ordinary contexts in which (L) is uttered. Again, the holist will hold that the membership conditions of a concept like *SHY* are non-holistic. The speaker who utters (L) is *not* requiring the believer to have a concept *SHY* that has *exactly the same inferential role* as the speaker's own shy-concept at t ; he simply wants the believer to have a token concept with the “standard shy-role” ([INTROVERTED, DOES NOT TALK MUCH...]). Again, this is compatible with the existence of a type concept SHY_H , whose membership conditions are indeed holistic but which just isn't part of the description D supplied by the speaker.

In passing, note how the “multiple types” solution gives us a way to charitably reinterpret the similarity strategy that was rejected in chapter 2 (sect. 3.2). The holist's strategy was to hold that concepts are holistically individuated, and that different thinkers never have the “same” concept but only “similar” ones. Now, if a token concept can belong to multiple types, then we should simply say that two token concepts had by different subjects will belong to many of the *same* types (e.g. the type CLARK), while also belonging to many *different* types (e.g. the type CLARK_H). Depending on what types two token concepts have in common, we will then say that these concepts are more or less “similar”. Of course, two tokens will share *infinitely* many types, since they share infinitely many properties. So if all types were equally important for

¹⁵¹ Objection: while a particular animal can belong to several sets of objects, it cannot belong to two species at the same time (nothing is both a cat and a dog, a cat and a tiger etc...). Kinds that are at the same “taxonomic level” (e.g. *cat* and *dog*) have mutually exclusive membership conditions. Reply: since the less specific CLARK will subsume the more specific CLARK_H , there is no reason to think that CLARK_H and CLARK will have mutually exclusive membership conditions. Similarly, it's perfectly possible for something to belong to both *mammal* and *cat*, since the former is located at a higher taxonomic level.

¹⁵² This point is made by Laurence and Margolis (1999, p. 76, fn. 30).

similarity, all concepts would be equally similar. For this reason, we need *context* to help us in our similarity ascriptions. In any given context, some types will be more salient than others: in the contexts described above, CLARK is more salient than CLARK_H, since it is part of the providing conditions imposed by the speaker. We will then say that two concepts are “similar” when they share enough contextually salient types (I will go back to this in ch. 5, sect. 4.3). In sum, the holist can account for publicity without having to invoke a mysterious notion of “primitive” similarity: the similarity between two concepts just reduces to their having certain (contextually salient) types in common. On its most charitable interpretation, then, the similarity strategy simply boils down to a version of the multiple types solution.

A possible objection against the appeal to multiple types is that it confuses *type concepts* with *classes* (or *sets*) of concepts¹⁵³. On Goodman’s original formulation of the type/token distinction, a token can only belong to one type at a time (Goodman 1968, p. 133). So a specific token of the word “dog” will only belong to one type, i.e. the type-word “dog”. On the other hand, an object can obviously belong to several classes of objects at the same time: indeed, it belongs to an *infinite* number of classes, since it instantiates infinite properties. So we should say that a concept belongs to multiple classes of concepts, but only one type concept¹⁵⁴. Of course, this would be inconsistent with the claim that one of my token concepts belongs to the type concept CLARK (to which one of your token concepts also belongs) *as well as* the holistic type concept CLARK_H (to which none of your token concepts belongs).

The non-Fregean holist might simply respond by changing his terminology. Instead of claiming that a token concept can belong to multiple types, he will now hold that a token concept can belong to multiple classes or sets of concepts: some of these classes (CLARK_H) have holistic membership conditions, while others (CLARK) do not. The objector might reply that this is not enough to satisfy the publicity requirement. What’s needed to account for the interpersonal applicability of (G5) is not simply that our token concepts belong to the same class of concepts CLARK; our token concepts must also belong to the same *type concept* CLARK (where this is a “proper” type concept, not just a class or a set). It is not clear to me that this reply would be warranted: some reasons should be provided for thinking that publicity requires sameness of (proper) type concepts. In any case, the non-Fregean does have a further response available; before turning to that response, however, let’s see how a Fregean holist might try to revise his view so as to accommodate publicity.

Revising holism (Fregean version): A Fregean holist might hold that only *some* abstract concepts are individuated by their global inferential role; other concepts are

¹⁵³ The objection is raised by Glock (2009, pp. 26-27) against Laurence and Margolis (1999).

¹⁵⁴ The reasons for formulating the distinction in this way are complex and cannot be discussed here. V. Dilworth (2003) for criticism, and Wetzel (2006) for an overview of the topic.

individuated less specifically, and it is *these* concepts that speakers refer to in providing and constraining contexts.

The required moves will of course parallel those made by the non-Fregean. Consider for instance a context like the above, in which CLARK is directly provided in the truth-conditions for (L). Our holist will hold that the provided concept CLARK is *not* individuated by its global inferential role at some specific time t for a specific subject S: it is simply individuated by its referent (Clark/Superman) and a part of its inferential role (which might e.g. correspond to the standard Clark-ish role). Different thinkers will then have CLARK as long as they have some mental representation referring to Clark/Superman and having the Clark-ish role, even if those representations differ in some of their inferential connections (I think Clark is handsome, you think he is not).

The holist can then maintain that, alongside CLARK, there exist many *other* concepts referring to Clark/Superman, and that some of these concepts *are* holistically individuated. For instance: there exists a concept CLARK_H individuated by its reference *plus* its global inferential role for subject S_1 at time t_1 (roughly speaking: by the totality of inferential dispositions/beliefs that S_1 has with respect to Superman/Clark at t_1 ¹⁵⁵); there exists a *different* concept CLARK_H^* individuated by its reference *plus* its global inferential role for subject S_2 ; and so on. Clearly, such concepts will *not* be shared. This, however, is unproblematic, since speakers will *not* provide such concepts in their ordinary belief ascriptions. In a standard context c , a speaker who utters (L) will presumably provide an “ordinary” concept of Clark, i.e. one that refers to Superman/Clark and is associated with those features that most people ascribe to him. There would be no need for the speaker to provide a more *specific* concept, e.g. one individuated by all his beliefs about Superman/Clark at the time of utterance, in order to communicate information about Lois’ mental state. Consequently, generalizations uttered in c will only require possession of the non-holistic CLARK, and not of the holistically individuated $\text{CLARK}_H/\text{CLARK}_H^*$. So, even if many concepts are holistically individuated, intentional generalizations uttered in providing contexts can still apply to multiple subjects. (Again, the same line will be run for concepts like SHY in constraining contexts; *v. supra* for more details).

We can now see how a non-Fregean might respond to the objection that sameness of (proper) type concepts is required for publicity. The response will follow closely the one just sketched in defense of Fregean holism. *Some* type concepts are not shared: these are type concepts such as $\text{CLARK}_H, \text{CLARK}_H^*, \dots$, which are holistically individuated by their global inferential role within a specific subject. For instance, my token concept C_1 , which instantiates the holistic type CLARK_H , will belong to a *different* type from your token concept C_2 , which might instantiate the holistic type CLARK_H^* . At the same time, however, each of us might have two *further* token concepts (call them, respectively, C_3 and C_4) which *do* instantiate the same type concept CLARK, just like two concrete

¹⁵⁵ Oversimplifying the Fregean picture, you can imagine CLARK_H as an extremely long description encoding all of S_1 ’s beliefs about Clark at t_1 .

tokenings of the same type-word “dog”. So the non-Fregean can still account for concept sharing, even if he grants that sameness of proper type concepts is required for this.

The two holistic responses I just described can be summarized in a single formula: even assuming contextualism, holism is compatible with interpersonal applicability since:

- (Non-Fregean): *some* type concepts have *non-holistic* membership conditions¹⁵⁶;
- (Fregean): *some* abstract concepts are *non-holistically* individuated.

Two problems can be noted with this strategy¹⁵⁷. First, the holist must now give up one of the two theses used to characterize his position in chapter 2 (sect. 2.3):

- **Concept individuation (IRS):** a concept is partially individuated by its inferential role: two concepts C_1 and C_2 are the same concept only if they have the same inferential role.
- **Holistic definition of inferential roles:** the inferential role of a concept C is the set of *all* the inferential connections in which C stands at a time t for a subject S .

The holist now maintains that only *some* concepts are individuated by their global inferential role: $CLARK_H$ is, but $CLARK$ is not; SHY_H is, but SHY is not¹⁵⁸. Once we also define “inferential role” as “the set of *all* the inferential connections in which C stands at a time t for a subject S ”, it follows that not all concepts are individuated by their inferential roles. Concepts like $CLARK$ and SHY are now individuated by only *part* of their roles (respectively, by [CALLED “ $CLARK$ ”, SHY , WEARS GLASSES...] and [INTROVERTED, DOES NOT TALK MUCH...]), not by *all* the inferential connections¹⁵⁹.

This is a substantive concession, but one that holists might be prepared to accept as long as *some* concepts are holistically individuated¹⁶⁰. However, even this more

¹⁵⁶ From now on, I will mostly ignore complications having to do with the “proper” type-token distinction (*v. supra*) and simply identify type concepts with classes/sets of token concepts, so as to conform to standard usage in the literature.

¹⁵⁷ In passing, note that the strategy would also have been available to the holist in response to the original Fodorian argument (ch. 2). However, we need a contextualist framework to implement the move effectively. Only within that framework can we see which “non-holistic” concepts will be contextually selected by speaker’s intentions, *how* they will be selected, and for *what purposes*.

¹⁵⁸ From now on, I will simply talk of “concept individuation” and leave it to the reader to reformulate my claims in terms that would be acceptable on each of our ontologies.

¹⁵⁹ Alternatively, the holist might decide to adopt a weaker definition of “inferential role”: the two options are substantially equivalent.

¹⁶⁰ Indeed, in at least one passage Block acknowledges explicitly that concepts can be typed in multiple ways, sometimes holistic and sometimes non-holistic:

moderate claim might now be challenged. To account for interpersonal applicability, the holist needs two assumptions:

- Some concepts are non-holistically individuated.
- In contexts where a generalization G is used, the speaker will make reference to “non-holistic” concepts rather than holistic ones.

The second assumption is of course as crucial as the first one: if speakers referred to holistic concepts, G would still fail to be interpersonally applicable. But that assumption raises a worry: *why* think that *any* concepts are holistically individuated, if there are *no contexts* in which speakers make reference to holistically individuated concepts?

The worry can be declined differently, depending on whether we are targeting a Fregean or a non-Fregean version of holism. Against a Fregean, one might argue as follows. To be warranted in postulating the *existence* of holistically individuated abstract concepts, we must have sufficient theoretical reasons. But suppose speakers never make reference to such concepts, so that they don’t play any role in our attitude reports or in the intentional explanations/predictions based on them; suppose we simply appeal to non-holistic concepts like CLARK or SHY for such purposes. Why, then, should we think that there *are* holistically individuated concepts in the first place?

The objection might seem less worrying for non-Fregean holists. After all, types are cheap: you can classify things any way you like without incurring any ontological commitments. You might then, if you so please, decide to type concepts into those that fall under the holistic type CLARK_H (only one) and those that don’t (all the others)¹⁶¹. Similarly, you might decide to classify things as “grue”/“non-grue” rather than “green”/“non-green”, or as “things-bigger-than-my-thumb-at- t ” and things that are not. Whether these classifications would be theoretically useful, of course, is a different matter. If there are *no contexts* in which we need to classify concepts in the very fine-grained way suggested by the holist, then the claim that concepts can be typed holistically will be true but uninteresting. What we have to show is that, given our purposes in certain contexts, we sometimes *need* to type concepts holistically. This is a much harder task, and one facing Fregean and non-Fregean holists alike.

This objection against the revised holistic view I sketched leads us to the next chapters. There, I will try to show (among other things) that there *are* some contexts in

Similar points about individuation of computations are familiar. We can type-identify computations by function computed, ignoring the specific algorithm deployed and its implementation, or by the algorithm, or by the implementation. We can type identify bathtubs architecturally - in terms of size, shape, weight, and decorative properties. Alternatively, we can type identify them economically, in terms of initial price, usable life, and thermal insulation, ignoring architectural features. Equivalence relations can be based on any of these features (Block 1993, p. 57)

¹⁶¹ Things will be different if we adopt a “proper” construal of types. In this case, the worry will arguably parallel the one arising for Fregean ontologies.

which holistically individuated concepts will play an important theoretical role. In these contexts, appealing to holistic concepts will turn out to be extremely useful, and perhaps indispensable. If so, then holism might still have a role to play in our best theory of concepts, even after we revise it substantially to make it compatible with the publicity principle.

4.2. Schneider's Appeal To Millianism

Before moving on to chapter 4, I would like to discuss an alternative strategy that the holist might employ in order to answer publicity objections without weakening his view. Susan Schneider¹⁶² has recently proposed a sophisticated holistic theory that is supposed to escape standard publicity worries. Having spelled out my favored theory of intentional generalizations, I am now in a better position to point out what I think is wrong with her view.

Schneider's central thesis is that LoT symbols (which she identifies with concepts) “[...] should be typed by their total computational roles, where, to a first approximation, the total computational role of a symbol is the role it plays in the relevant ‘program’” (Schneider 2011, p. 120). (Roles are then defined as including inferential links to other concepts, as well as links to perceptual inputs/behavioral outputs). Schneider agrees that different subjects will never have concepts of the same type on her view of concept-individuation, but she doesn't take this to be incompatible with the interpersonal applicability of intentional generalizations and explanations, for two reasons.

First, like Block¹⁶³, Schneider notes that several psychological generalizations do not make reference to *specific* propositional attitudes. One instance of this are the “fundamental” generalizations we discussed in chapter 2: “For any action *a* and any goal *g*, if one wants *g* and also believes that *a* is required for *g*, then one will try to do *a*” (Block 1998). Other “unspecific” generalizations are offered by cognitive psychology: consider for instance “George Miller's generalization about the upper limit on the number of items in working memory: the magical number seven, plus or minus two” or “‘syntactic’ or proof theoretic versions of logical generalizations like *modus ponens* and conjunction elimination” (Schneider 2011, p. 144). Finally, some important methods of explanation in cognitive science will also avoid referring to specific attitudes: v. for instance the method of “functional decomposition”, which explains the functioning of a certain cognitive capacity (e.g. working memory) by decomposing it in terms of its more basic components and then describing their functioning without making reference to any *specific* symbols in the process (Schneider 2011, pp. 140-42).

Since the generalizations/explanations just described do not make reference to specific attitudes, *a fortiori* they do not require the subjects they cover to have any specific concepts in common. As noted in chapter 2, however, this is not enough as a

¹⁶² Schneider 2005; 2009 a,b; 2011.

¹⁶³ V. ch. 2 (sect. 3.1).

defense of holism (v. sect. 3.1, problems 2-4): there *still is* a large class of intentional generalizations, explanations and ascriptions which *do* make reference to specific attitudes, such as:

G2) If a subject S wants to get water and believes that if she opens the fridge she will get water, then other things being equal S will open the fridge.

How will the holist account for the interpersonal applicability of these generalizations? Schneider's response is that attitude ascriptions, and the generalizations/explanations embedding them, have a *Millian semantics*¹⁶⁴. Since ascriptions are Millian, they don't encode any information about the *concepts* under which subjects believe the Russellian propositions expressed by "that" clauses. Consequently, our schematic ascription "S believes that if she opens the fridge she will get water" will be true of two subjects S₁-S₂ as long as they both believe the relevant Russellian proposition *under some concept or other*. S₁-S₂ are therefore *not* required to believe that proposition under some specific concept WATER. So, even though holism entails that WATER cannot be had by more than one subject, this is compatible with the ascription being true of both S₁-S₂. Consequently, generalizations and explanations embedding that ascription will be interpersonally applicable under holism.

Schneider's defense of holism is more sophisticated than Block's: by appealing to a Millian semantics, the holist can account for the interpersonal applicability of those generalizations that make reference to specific attitudes, while still holding that concepts aren't shareable. The holist can then reject the publicity principle (in both its original and its contextualist version) and avoid weakening her view in the way I suggested previously. There is no need to concede that *some* concepts must be shareable, and thus non-holistically individuated, since concepts are simply not involved in our intentional generalizations: all that's required for a generalization to apply to a group of subjects is that they believe the relevant Russellian proposition *under some concept or other*.

Now, in chapter 1 I offered several reasons to reject the Millian semantics on which Schneider's defense hinges. Some of those arguments concerned our intuitions about ascriptions in general, while others were focused on intentional generalizations and explanations. In this section, I will concentrate on the latter class of arguments in order to highlight the problem with Schneider's strategy. Once we adopt a Millian semantics, we "impoverish" the content of ascriptions, which will now encode less information about the belief states of the relevant subjects. In particular, no information will be communicated about the concepts under which those subjects believe the relevant propositions. This has the advantage of making the conditions for interpersonal applicability easier to satisfy: in particular, it is now possible for different subjects to satisfy those conditions even if they have *no concepts* in common, as the holist

¹⁶⁴ In her terminology: they are "broad". V. Schneider (2005) and Schneider (2011, pp. 136-140).

maintains. But adopting a Millian semantics also has its costs. By impoverishing the content of generalizations, we make them *less useful* for purposes of *explanation* and *prediction*: wider applicability is bought at the expense of explanatory and predictive power. The holist's appeal to Millianism makes it impossible to account for the fact that certain generalizations are routinely used to explain/predict intentional behavior; only a non-Millian semantics can account for this. On that semantics, however, specific concepts will again be involved in our generalizations: Schneider's defense of holism will then become unavailable, and the holist will have to adopt the weakened view described in the previous section.

To see why Schneider's appeal to Millianism runs into trouble, I want to analyze two close relatives of the generalizations examined in our discussion of Braun's view (v. (12)-(13), ch. 1, sect. 4). Consider first:

G6) If a subject S believes that Superman is nearby and believes that Lex Luthor is nearby, then other things being equal S will run towards Superman¹⁶⁵.

(G6) appears true; moreover, it seems we can use it to explain and predict behavior¹⁶⁶. For familiar reasons, however, (G6) will face several apparent counterexamples once we assume a Millian semantics. Consider Lucy, who:

- a) Does not accept the identity claim “Superman is Clark”.
- b) Sincerely utters “Clark is nearby”.
- c) Sincerely utters “Luthor is nearby”.

Clearly, Lucy believes that Luthor is nearby; moreover, she also believes that Clark is nearby. Assuming Millianism, it follows that she also believes that *Superman* is nearby. So, assuming Millianism, Lucy seems to satisfy (G6)'s antecedent. But Lucy might perfectly well *not* run towards Superman in the circumstances described above (suppose Superman/Clark is dressed in his reporter clothes, so that Lucy would *not* accept “Superman is nearby”). Therefore, Lucy seems to satisfy (G6)'s antecedent but not its consequent, thus constituting an apparent counterexample to (G6). Moreover, similar counterexamples can be multiplied at will (imagine a crowd of people in Lucy's predicament who run away screaming instead of running towards Superman/Clark).

Of course, Schneider's reply would be that Lucy *does not* in fact satisfy (G6)'s antecedent: other things are *not* equal in her case, since she represents Superman/Clark in mismatching ways¹⁶⁷. As usual, I will grant for the sake of the argument that the Millian is allowed to include a “matching ways” requirement in the *ceteris paribus*

¹⁶⁵ I borrow (G6) from Aydede (1998) and Schneider (2011, p. 149), with some modifications.

¹⁶⁶ In what follows I will mostly focus on explanation, although everything I say would also apply to prediction.

¹⁶⁷ Cf. ch. 1 (sect. 4) for more details about the *ceteris paribus* strategy.

clause of generalizations like (G6)¹⁶⁸. Even then, the move is not powerful enough to save a Millian semantics for generalizations, since it cannot be applied to *negative* ones. Consider:

G7) If a subject S **does not believe** that Superman is nearby and believes that Lex Luthor is nearby, then other things being equal S will **run away**.

Like (G6), (G7) appears true. Moreover, it seems it can be used for purposes of explanation and prediction; indeed, (G7) seems to be precisely the generalization we would use to explain why Lucy runs away in the circumstances described above. Of course, however, Millianism entails that we *cannot* use (G7) to explain Lucy's behavior, since she doesn't satisfy the antecedent (by assumption, she *does* believe Clark is nearby). But since (G7) can be so used, Millianism is false. To put the same point differently: if Millianism was true, the following DN argument would be an *unsound* explanation of Lucy's behavior (premise 2 would be false):

- 1) If a subject S does not believe that Superman is nearby and believes that Lex Luthor is nearby, then other things being equal S will run away.
- 2) Lucy does not believe that Superman is nearby.
- 3) Lucy believes that Lex Luthor is nearby.
- 4) Other things are equal.
- 5) Lucy will run away.

But (1)-(5) certainly seems to be a sound explanation of why Lucy ran away in the above circumstances!

As with the positive generalization (G6), Millianism seems to make negative generalizations like (G7) useless for purposes of intentional explanation. This time, however, the *ceteris paribus* move will not help: the problem with (G7) is that Lucy doesn't satisfy the antecedent, *not* that she risks satisfying the antecedent but not the consequent (as with (G6))¹⁶⁹. If (G7) can be used to explain/predict Lucy's behavior, it cannot have a Millian semantics.

On the contrary, our non-Millian contextualist semantics can easily account for both the positive (G6) and the negative (G7). On that semantics, in all contexts in which (G6) appears true Lucy will *not* satisfy the antecedent. In such contexts, “S believes that Superman is nearby” will only be true of subjects who believe <Superman, is nearby>

¹⁶⁸ Schneider provides several reasons to include this requirement: v. Schneider (2005) and Schneider (2011, ch. 8).

¹⁶⁹ Alternative moves available to the Millian would also not help: v. ch. 1 (sect. 4) for more details.

under SUPERMAN, while Lucy only believes that proposition under CLARK. So Lucy is not a counterexample, and we have no need to make any controversial assumptions about (G6)'s *ceteris paribus* clause. Moreover, our semantics can also account for (G7): in all contexts in which we can use (G7) to explain Lucy's behavior, she *does* satisfy the antecedent. In these contexts, "S believes that Superman is nearby" is *false* of Lucy, since, again, she believes <Superman, is nearby> under CLARK and not under SUPERMAN. Therefore, "S does not believe that Superman is nearby" is true of Lucy, the antecedent is satisfied and (G7) can be used to produce a sound DN explanation of her behavior.

In conclusion: if intentional generalizations have a Millian semantics, we will have *troubles* accounting for our use of *positive* generalizations like (G6) in explanations/predictions; moreover, we will *certainly* be unable to account for our use of *negative* generalizations like (G7). But recall the central thesis in Schneider's strategy: holism is compatible with the interpersonal applicability of intentional generalizations, since such generalizations have a Millian semantics. My arguments in this section show that strategy to be unsuccessful: intentional generalizations cannot have a Millian semantics, or we would not be able to use them to explain/predict the behavior of subjects like Lucy.

Recall, moreover, that a non-Millian contextualist semantics can straightforwardly account for both (G6) and (G7). On that semantics, however, there will be many contexts in which a generalization only applies to a group of subjects if these subjects share a specific concept C: this is precisely the (new) publicity principle discussed in section 3. For instance, our semantics entails that a generalization like (G6) will only apply to subjects who have the concept SUPERMAN. This is straightforwardly incompatible with Schneider's holistic position: on that view, the concept SUPERMAN can only be had by one subject at a time, so (G6) will not cover multiple subjects. On our best semantics for generalizations, holism is still incompatible with their interpersonal applicability.

In an attempt to defend her strong holistic position through Millianism, Schneider makes the content of intentional generalizations too impoverished to capture their explanatory and predictive power. Only a non-Millian, contextualist view can account for the fact that we successfully employ generalizations like (G6)-(G7) in our folk-psychological practices. On that view, however, specific concepts are often involved in our generalizations, and concept publicity must be respected. Schneider's "strong" holism thus becomes untenable, and the only option left for the holist is to adopt the "weakened" view I sketched in the previous section.

Chapter 4

Publicity And The Fregean Constraint

1. Introduction

In the present chapter (ch. 4) and the next (ch. 5) I will tackle our main problem, that of establishing a set of conditions for concept individuation/possession, from a slightly different angle. I will discuss the relationship between the publicity principle and a second widely endorsed constraint on a theory of concepts. Publicity holds that subjects who are covered by the same generalizations must have the *same* concepts; our second constraint, which I will call “the Fregean constraint”, pulls in the opposite direction. To a first approximation, it holds that a subject who finds himself in a certain kind of scenario (a so-called “Frege case”) must have *distinct* coreferential concepts for one and the same object or property. For instance, someone like Lois Lane must have two different concepts SUPERMAN and CLARK KENT for one and the same person, Superman/Clark.

The Fregean constraint has also been widely recognized as an important *desideratum* for a theory of concepts, although it isn’t always openly stated¹⁷⁰. As we will see later, however, many authors who don’t endorse the constraint explicitly still seem to conform to it when developing their theories, while others have accepted principles in the vicinity that have very similar consequences¹⁷¹. Getting clearer on the implications that the Fregean constraint will have for a theory of concepts is thus of paramount importance: this is what the next two chapters are about.

To a first approximation, my main thesis will be that there is a tension between publicity and the Fregean constraint: theories of concepts that satisfy the latter seem to run into trouble with the former¹⁷². My aim, however, is not to show that the two constraints are inconsistent; on the contrary, I believe they can be made compatible by adopting various strategies, which I will examine in the next chapter. I rather aim at showing that, *once* we adopt these strategies, certain important consequences for a theory of concept individuation/possession will follow. I am particularly interested in two of these consequences, since they bear directly on the issues discussed in chapters 2-3. Examining the relationship between our two constraints will help us show that:

- The “modified” version of *holism* sketched at the end of chapter 3 (sect. 4.1) will have a useful role to play in our best theory of concepts.
- The most plausible version of the publicity principle is the *contextualist* one developed in chapter 3.

¹⁷⁰ The constraint does have its opponents: v. especially Millikan (2000, chs. 9-12). V. also Sainsbury and Tye (2012), whose position will be discussed later.

¹⁷¹ Cf. sect. 2 *infra*.

¹⁷² To my knowledge, no one in the literature has stated the problem in these terms, although issues in the vicinity have certainly been recognized: v. especially Aydede (1998; 2000 a,b), Rupert (2008), Schneider (2009 a,b; 2011).

Let me say a bit more about the second goal. I want to convince you that, if you like the publicity principle, then you should endorse the contextualist picture sketched in chapter 3; assuming the publicity principle is true, we should hold a contextualist version of it. (This is because, as we will see later, a non-contextualist version would make it hard for us to solve the tension between publicity and the Fregean constraint).

Because I want the next chapters to provide independent support for my version of publicity, I will *not* rely on that version of the principle in what follows. I will instead assume (something like) the original publicity principle defended by Fodor and others¹⁷³ and try to convince you that, once we assume that principle, we will have to accept a contextualist version of it. In addition, the publicity principle I'll presuppose in chapters 4-5 will be *more general* than the one discussed in chapters 2-3. That principle was exclusively concerned with psychological generalizations, while my focus in the next chapters will be broader: instead of concentrating on generalizations alone, I will be taking all the “standard” arguments for the publicity of concepts at face value. As you might recall from chapter 2 (sect. 2.2), those arguments were:

- It seems plausible to hold that ordinary thinkers can have the same desires, beliefs, intentions etc.... But how could two thinkers have the *same propositional attitude*, unless they also shared the corresponding concepts? For instance: how could two thinkers both have the belief that dogs bark, unless they both had the concept DOG?
- In order for *communication* to be successful, it should be possible for two communicating subjects to be in the same mental state. If I say “dogs bark”, our communicative exchange will be successful only if I get you to entertain the same thought DOGS BARK which I have in mind and intend to communicate to you. But how could we entertain that same thought if we didn't both have DOG?
- In order for me and you to *genuinely agree* on a certain subject matter we must be in the same mental state: for instance, we won't agree on whether dogs bark unless we both accept the thought that dogs bark. But how could we accept that same thought unless we both had its constituting concepts? (*Mutatis mutandis* for genuine *disagreement*).

If sound, these arguments establish (together with the one from psychological generalizations) that concepts are routinely shared by ordinary thinkers for a variety of purposes:

- Whenever two thinkers have the same propositional attitude, there will be a set of concepts involved in that attitude which are shared by both thinkers.

¹⁷³ Cf. ch. 2 (section 2.2).

- Whenever a communicative exchange succeeds, the two parties must share all those concepts that constitute the communicated thought.
- Whenever two parties genuinely agree on a certain subject matter, they must share all those concepts which constitute the thought that was agreed upon.
- Finally, whenever two subjects are covered by the same generalization they must share all the concepts involved in that generalization.

This broader formulation of publicity is the one I will be presupposing in the following chapters. Put more concisely, we could state it as follows:

(PUB) (general version): For any two subjects S_1-S_2 who [share an attitude/communicate successfully/genuinely agree or disagree/ are covered by the same generalization], there is a set of concepts $[C_1, C_2, \dots C_n]$ such that S_1-S_2 would *not* [share an attitude/communicate successfully/genuinely agree or disagree/ be covered by the same generalization] *unless* they both had $[C_1, C_2, \dots C_n]$.

In less formal terms: ordinary intentional agents often have many of the *same* concepts, and this plays an essential role not only in their being covered by the same psychological laws, but also in a number of other aspects of their cognitive lives. This formulation of publicity is somewhat less precise and rigorous than the various versions that were discussed in previous chapters, but this should not be a problem; a fairly general and intuitive formulation will be enough to raise the problem we are interested in.

Also, notice that much of our previous discussion focused on the *motivations* for publicity: I discussed standard arguments by Fodor and others, argued that it's not clear whether they are valid, and then offered an alternative argument based on our best semantics for ascriptions. In the next chapters, I will focus instead on the *implications* of publicity. In particular, I will try to determine whether our constraint is inconsistent with an independently plausible principle, the Fregean constraint. That might easily tempt us to reject publicity: fortunately, I will show that this is not necessary and that the two constraints can coexist. My hope is that this will make publicity more attractive and easier to accept, especially when declined in the contextualist version I favor.

Now that I have clarified what version of publicity I will be assuming, here is how I propose to proceed. Having described our second constraint and the motivations behind it (sect. 2), I will move on to consider its relationship with publicity. I will examine some of the main theories of concept individuation/possession currently on the market and show that they all have a similar structure (sect. 3). Then (sect. 4), I will show that theories with that kind of structure have problematic consequences: they *either* fail to satisfy the Fregean constraint *or* turn out to be inconsistent with publicity. In the next

chapter, I will explain what I take the source of the problem to be, and how I think it should be solved. My ultimate goal will be to show that:

- On all possible solutions, *holism* will turn out to play an important role in our best theory of concepts.
- On at least one of those solutions, a *contextualist* version of the publicity principle becomes extremely plausible.

I will then conclude by drawing a general picture of concept individuation and possession, which, I think, is strongly supported by the results of our discussion throughout chapters 1-5.

2. The Fregean Constraint

2.1. Frege Cases And Rationality

In a number of cases, a good theory of concepts should count certain subjects as having *distinct* concepts for one and the same object or property. Following standard usage, I will refer to such cases as “Frege cases”¹⁷⁴. Consider our familiar scenario involving Lois Lane and Superman/Clark. Lois, we might suppose, is a normal subject under all relevant respects: she is intelligent, attentive, reasonably cautious in forming her judgments and so on. She has certain evidence (perceptual, testimonial etc...) regarding the identity of the person called “Superman” and the person called “Clark”. On the basis of that evidence, she has formed certain beliefs about the matter, and as a result she sincerely accepts the claims “Superman can fly” and “Clark cannot fly”. Clearly, Lois’ case is perfectly possible. At the same time, however, it also presents a number of puzzling features that call for an explanation. Here is one that will be particularly important for our purposes:

- Since Lois accepts “Superman can fly”, she seems to believe *of* Superman/Clark that he can fly; and, since she accepts “Clark cannot fly”, she also seems to believe *of* Superman/Clark that he *cannot* fly. So Lois believes, *of* one and the same individual, *both* that he can fly *and* that he cannot fly¹⁷⁵. That is, Lois is ascribing contradictory properties to one and the same object, even though she knows full well that nothing can be both F and not-F at the same time.
- Still, Lois doesn’t seem *irrational* in accepting the above claims. Given her evidence, she has excellent reasons to take them to be true, and little or no reasons to doubt either of them. After all, she has always seen the person called “Superman” fly while wearing his red and blue outfit, not while wearing his reporter uniform; she has all reasons to believe that the person called “Clark Kent” is a normal human being in all relevant respects; and so on¹⁷⁶. Given her epistemic situation, then, Lois certainly doesn’t seem guilty of irrationality in (mistakenly) accepting both of the above claims.

¹⁷⁴ V. e.g. Aydede and Robbins (2001), Schneider (2005), Rupert (2008), who use the same terminology.

¹⁷⁵ Notice that Millians and non-Millians *agree* that subjects in Frege cases have contradictory *de re* beliefs. For instance, it is usually agreed that Lois believes, *of* Superman/Clark, that he can fly and that he cannot fly; the disagreement is on whether she also believes *that* Clark can fly. V. Schiffer (2006, pp. 361-363) and Salmon (2006).

¹⁷⁶ Of course, the story can be modified so that Lois’ evidence against the identity of Superman and Clark is almost irresistible.

- But *how* can Lois *rationally* believe, of one and the same individual, that he both can and cannot fly? This appears especially puzzling since, again, Lois wholeheartedly *agrees* that nothing can be both F and not-F at the same time!

A satisfactory theory of the mind should be able to explain why Lois is rational despite her contradictory beliefs. Clearly, subjects in other popular “Frege cases” will also raise the same problem. For instance: in Kripke’s (1979) famous case, Peter believes of Paderewski that he has musical talent (in some contexts, he accepts “Paderewski has musical talent”), but at the same time believes, of the very same person, that he does not have musical talent (in other contexts, he accepts “Paderewski does not have musical talent”); Oedipus believes, of Jocasta, that she is and she is not his mother; and so on. Given their evidence, these subjects don’t seem irrational in their doxastic behavior, and yet they all form contradictory *de re* beliefs on the basis of that evidence. The general problem posed by Frege cases can then be stated thus:

The Rationality Problem: *Why* are Lois, Peter and other subjects in Frege cases *rational*, even though they all believe, of the same object x , that x is both F and not-F¹⁷⁷?

2.2. Explaining Rationality: The Fregean Constraint

To solve the Rationality Problem, we must explain why Lois and other “unenlightened” subjects are rational despite their contradictory beliefs. Different explanations have been proposed within RTM, but I believe it’s possible to isolate a basic strategy which would be endorsed by most RTM theorists. More specifically, there is a basic assumption made by all these theorists about how to explain the rationality of characters like Lois. This assumption, which I will call “the Fregean Constraint”, is our second constraint on a theory of concepts¹⁷⁸.

To see what the Fregean constraint amounts to, let’s start by analyzing some specific explanations that have been proposed by RTM theorists to account for the rationality of unenlightened thinkers like Lois. (Some of them will be familiar from ch. 1, but it will be useful to repeat them here). Here is the account proposed by Braun within his LoT-RTM framework:

[Lois] would not be able to deduce any contradiction from [the propositions expressed by (1) and (2n)¹⁷⁹], given the ways in which she believes them, for no contradictory sentence can be validly

¹⁷⁷ My way of stating the problem is partially based on the way Braun presents the Substitution objection against Millianism (ch. 1, sect. 2.1). The relationship between the two constitutes an interesting issue in itself, but I cannot discuss this here.

¹⁷⁸ Some of these theorists have not explicitly acknowledged the assumption, but the explanations they propose do seem to rely on it. Braun (1998) is a case in point: v. next paragraph.

¹⁷⁹ (1) = “Superman can fly”; (2n) = “Clark cannot fly”.

derived (in the syntactic sense) from sentences (1) and (2n) alone. Thus it may be no more irrational for her to have (1) and (2n) in her belief-box than for her to have ‘Gingrich is a Republican’ and ‘Clinton is not a Republican’ in her belief-box (Braun 1998, pp. 575-576).

And here is how Recanati summarizes the explanatory strategy that was originally suggested by Frege and afterwards employed by many others (including himself):

In the Fregean framework, modes of presentation provide a solution to the following puzzle: A rational subject can think of a given object *a* both that it is and that it is not *F* – how can that be? Frege solved the problem by appealing to modes of presentation over and above the objects thought about. A rational subject can believe of *a*, thought of under a mode of presentation *m*, that it is *F*, and at the same time believe of the same object *a*, thought of under a different mode of presentation *m'*, that it is not *F*. Insofar as the modes of presentation are distinct, there is no irrationality (Recanati 2009, pp. 254-255).

Let’s try to provide a more general formulation of the strategy suggested by these passages. (Some RTM theorists might disagree about the details here, but I have no ambition to offer a faithful reconstruction of a particular philosophical view: I am only interested in what is in common between the various specific explanations that have been proposed)¹⁸⁰:

The Standard Explanation

- i. Lois has two *distinct* concepts SUPERMAN and CLARK, both of which refer to Superman/Clark.
- ii. Since concepts are the constituents of thoughts, Lois’s thoughts SUPERMAN CAN FLY and CLARK CAN FLY are *distinct* thoughts [from premise (i)].
- iii. Lois stands in the belief-relation to (from now on: “accepts”) the thought SUPERMAN CAN FLY. She also accepts the thought CLARK CANNOT FLY, which is the *negation* of the thought CLARK CAN FLY. So Lois believes, of Superman/Clark, that he can fly and that he cannot fly in virtue of accepting SUPERMAN CAN FLY and the negation of the thought CLARK CAN FLY.
- iv. Therefore, Lois believes, of Superman/Clark, that he can fly and that he cannot fly in virtue of accepting a thought T_1 and the negation of a thought T_2 *distinct* from T_1 [from (ii), (iii)].
- v. If someone believes, of one and the same object, that it is both *F* and not-*F* in virtue of accepting a thought T_1 and the negation of a thought T_2 *distinct* from T_1 , then that person is not irrational¹⁸¹.

¹⁸⁰ For further instances of the same strategy, v. e.g. Salmon (1986, chs. 7-8) and Schiffer (1990, pp. 251-252). V. Millikan (2000, chs. 9-10) for a similar reconstruction of the standard explanation. (Notice that Braun and Recanati respectively identify modes of presentation with LoT symbols and mental files, so they would both agree on restating their explanation in terms of “concepts” as I do in this section).

¹⁸¹ Other things being equal: there might be other reasons why the person is irrational, but she won’t be irrational in virtue of accepting T_1 and the negation of a distinct thought T_2 .

vi. Therefore, Lois is not irrational [from (iv), (v)].

(A proponent of the Standard Explanation will then hold that Lois *would* be irrational if she believed of Superman/Clark that he can and cannot fly in virtue of accepting a thought (e.g. SUPERMAN CAN FLY) *and* the negation of that very *same* thought (e.g. SUPERMAN CANNOT FLY). Since Lois accepts and rejects¹⁸² distinct thoughts, however, she is not irrational in her attitudes).

A lot could be said about the Standard Explanation: in particular, one might want an argument for the crucial principle (v) connecting rationality and sameness/difference in thought. A discussion of this point would be of great interest, but it is not something I will attempt here. In what follows, I will simply *assume* that the Standard Explanation (or something close to it) is a sound explanation of Lois's rationality: I am interested in drawing the implications that the explanation has for our theory of concepts, not in assessing its merits¹⁸³. In any case, the intuitive idea behind it should be clear enough. There is nothing irrational in accepting a thought and the negation of a different thought, even if, by doing so, one ends up ascribing contradictory properties to the same object. *Because* the two thoughts are distinct, the thinker is not in a “position to see” that, by accepting them, he holds contradictory *de re* beliefs about the same object. If the two thoughts were the same thought, on the other hand, he would be able to recognize that his beliefs are contradictory. If he then persisted in accepting both thoughts, he would be knowingly ascribing contradictory properties to the same object, which arguably constitutes an irrational doxastic behavior.

Crucially, once we endorse the Standard Explanation as a correct account of Lois' rationality, we also seem to be committed to a certain general principle concerning the identity conditions of concepts. This principle is our second constraint on a theory of concepts; I will refer to it as “the Fregean constraint”¹⁸⁴. The principle has been stated in a variety of different ways; my formulation is in fact a mere terminological variant on some better known versions (I prefer my formulation because it fits better with the terms of our discussion in the following chapters).

Let's say that, when a subject has a *de re* belief about some *x* in virtue of accepting a thought that is partially constituted by some concept *C*, then *C* is “involved” in the subject's belief. For instance, we will say that Lois' SUPERMAN-CLARK are “involved” in her contradictory beliefs about Superman/Clark: Lois believes of Superman/Clark that he can fly in virtue of accepting SUPERMAN CAN FLY, and she believes of Superman/Clark that he cannot fly in virtue of accepting CLARK CANNOT FLY. We can then state our constraint as follows:

¹⁸² Rejecting a thought simply consists in standing in the belief-relation to the negation of that thought.

¹⁸³ V. Millikan (2000) for criticism.

¹⁸⁴ I borrow the phrase from Schiffer (1990, p. 252).

The Fregean Constraint (FC): If two concepts C_1-C_2 are involved in the contradictory beliefs of a *rational* subject S at time t about some object x , then C_1-C_2 are *not the same concept*¹⁸⁵.

It's easy to see that, once we accept the Standard Explanation as a sound explanation of Lois' rationality, we will also have strong reasons to accept (FC). There are no significant differences between Lois' case and other Frege cases in which someone rationally holds contradictory beliefs (cf. for instance Peter or Oedipus)¹⁸⁶. So, if the Standard Explanation constitutes a sound explanation of Lois' rationality, it should also explain why these other subjects are rational. But then, whenever a rational subject has contradictory beliefs about some x , we must assume that he does so in virtue of accepting two *distinct* thoughts T_1-T_2 constituted by *distinct* concepts C_1-C_2 . This, of course, is exactly what (FC) amounts to: if two concepts are involved in the contradictory beliefs of a rational subject, then they cannot be the same concept, or we would not be able to use the Standard Explanation to explain the subject's rationality.

(FC) is our second constraint on a theory of concepts. Before we start analyzing its relation with publicity, I should note that the rationality problem is not the only puzzle raised by Frege cases: the *behavior* of subjects involved in such cases and the *cognitive significance* that certain claims will have for them also seem to call for an explanation. Consider Lois again:

¹⁸⁵ Several authors have (more or less explicitly) accepted some version of (FC): v. among others Salmon (1986), Braun (1998), Recanati (2009). My way of stating (FC) is inspired by Schiffer's version of the principle:

[...] Frege's constraint has two parts. First it says that a rational person x may both believe and disbelieve that a certain thing or property y is such and such only if there are distinct modes of presentation m and m' such that x believes y to be such and such under m and disbelieves it to be such and such under m' . Then it says that there are distinct modes of presentation m and m' such that x believes y to be such and such under m and disbelieves it to be such and such under m' only if x fails to realize that m and m' are modes of presentation of one and the same thing. In other words, you can't rationally believe and disbelieve something under one and the same mode of presentation, or under modes of presentation which you realize are modes of presentation of one and the same thing (Schiffer 1990, p. 252).

(V. also Schiffer 1987, 2006). As Recanati points out in the passage quoted above, the idea behind (FC) is an old one: it goes back to Frege's idea of postulating distinct modes of presentation to explain the cognitive significance of certain identity claims (v. Frege 1956). For an illuminating discussion of the Fregean appeal to modes of presentation, v. Millikan (2000); this chapter was partly inspired by Millikan's penetrating remarks on this issue.

¹⁸⁶ This premise might be questioned; for instance, it would be rejected by Sainsbury and Tye (v. sect. 4.2 *infra* for discussion). Here I'm not interested in offering a bullet-proof argument for (FC), only in spelling out the motivations behind it.

- Lois will go out with Superman/Clark if he asks her “Would you like to go out for dinner?” while wearing his superhero outfit, but not if he asks the same question while wearing his reporter clothes.
- Lois will find the claim “Superman is Clark Kent” informative, but she will find the claims “Superman is Superman” and “Clark Kent is Clark Kent” uninformative.

Why does Lois behave in this way, if the person asking her out is the same in the two cases? And why does she find some identity claims about Superman/Clark informative and others uninformative, if they all ascribe exactly the same property (*being identical to himself*) to the very same guy?

A lot would have to be said about these questions and their relation to the rationality problem. What matters for our purposes, however, is that many RTM theorists have explained these features of Frege cases by taking the subjects involved to have *distinct* coreferential concepts. For instance, many would explain Lois’ behavior by claiming that Lois stands in the desire-relation with the thought GOING OUT WITH SUPERMAN, but not with the *distinct* thought GOING OUT WITH CLARK KENT¹⁸⁷. And they would explain the different informativeness that identity claims have for her by taking Lois to associate the uninformative claims with the thoughts SUPERMAN IS SUPERMAN/CLARK KENT IS CLARK KENT (which she accepts as self-evident), and the informative claims with the *distinct* thought SUPERMAN IS CLARK KENT (which she won’t accept until becoming enlightened about Clark’s secret identity)¹⁸⁸.

Of course, these accounts of behavior and cognitive significance assume that Lois’s SUPERMAN-CLARK KENT are *distinct* concepts: *mutatis mutandis* for other Frege cases. Besides the rationality problem, then, RTM theorists will have further reasons to ascribe

¹⁸⁷ Cf. for instance Fodor:

If beliefs (and the like) are relations to syntactically structured mental representations, there are indeed two parameters of belief individuation, just as Frege requires: Morning Star beliefs have the same conditions of semantic evaluation as Evening Star beliefs, but they implicate the tokening of different syntactic objects and are therefore different beliefs with *different causal powers* [...]. That believing P and believing Q may have different causal powers even if ‘P’ and ‘Q’ have the same semantic value shows up in all those operas where the soprano dies of mistaken identity (Fodor 1998, p. 39, my emphasis).

¹⁸⁸ Cf. for instance Peacocke, who defines concept sameness and distinctness in terms of cognitive significance:

Distinctness of Concepts: Concepts C and D are distinct iff there are two complete propositional contents that differ at most in that one contains C substituted in one or more places for D, and one of which is potentially informative while the other is not (Peacocke 1992, p. 2).

distinct coreferential concepts to the subjects involved in a Frege case, since this will often be required by their accounts of behavior and cognitive significance¹⁸⁹.

¹⁸⁹ Cf. for instance Fodor (1994, 1998, 2008), who seems to be interested in problems of behavior and cognitive significance more than in the rationality problem. V. also Prinz (2002).

3. Individuation Criteria For Concepts

3.1. Some Criteria

Now that our two constraints are on the table, it's time to start examining the relation between them. I will consider some important theories of concepts that have been proposed in the recent literature and note that they all seem to present a problem. These theories put forward sets of "sameness conditions" (or "individuation criteria") for concepts, where these conditions are of a specific form (they are what I will call "unrestricted" conditions). In doing so, the goal of many of these theories is to satisfy one or both of our constraints. As we will see in section 4.2, however, all the unrestricted criteria that are powerful enough to satisfy (FC) also seem to be *inconsistent* with (PUB). This is a problem: if our two constraints are consistent, we should be able to find an individuation criterion that can satisfy both. I will then consider a possible way to "patch" some of these criteria so as to make them compatible with publicity (sect. 4.3). As we'll see, this solution still entails that several concepts are *not* shared; whether this consequence is compatible with the publicity principle is something I will discuss in the next chapter.

Much of our discussion will concern "individuation criteria" for concepts that are "unrestricted". In order to introduce these notions, it will be useful to look at some concrete examples first. We'll start with the "originalist" theory of concepts recently proposed by Sainsbury and Tye¹⁹⁰ (2011, 2012):

Originalism answers the question: what are the necessary and sufficient conditions for the concept C_1 to be the same concept as the concept C_2 ? According to originalism, every concept has exactly one originating use, and every originating use of a concept is an originating use of just one concept. Hence we can offer the following necessary and sufficient condition for concepts to be the same:

(O) Concept $C_1 = \text{concept } C_2$ iff the originating use of $C_1 =$ the originating use of C_2 .

In this picture, each use U of a concept is a use of the unique concept that lies at the origin of the R-linked chain of uses to which U belongs (Sainsbury and Tye 2011, p. 105).

We will say more about origins later. For now, what matters is that ST are proposing an "individuation criterion" for concepts (a set of necessary and sufficient conditions for concept identity) which holds "unrestrictedly" (it applies to all concepts).

ST are also explicit in taking publicity to be an important constraint on a theory of concepts¹⁹¹:

¹⁹⁰ From now on, "ST".

¹⁹¹ Indeed, this is supposed to be one of the main motivations behind originalism: v. sect. 4.2 *infra* for more details.

Concepts are public and shared. No doubt we and you think of Austin in somewhat different ways. But we all have the same concept AUSTIN, shown by the sameness of some of our thoughts, and logical relations among others. If you and we think that Austin is a cool city, we think the same thought, and hence share the concepts that constitute it, and hence share the concept AUSTIN. If you think that Austin is the music capital of the world and we disagree, we think a thought that conflicts with yours, and this is again explicable in terms of our sharing the concept AUSTIN. This is consistent with our associating the concept with different information, which is what makes us say that we think of Austin in different ways (Sainsbury and Tye 2012, p. 21).

ST also think subjects involved in Frege cases must have distinct coreferential concepts in order to account for puzzles of cognitive significance:

[...] the pattern exemplified by the thought that Hesperus is Hesperus is distinct from the pattern exemplified by the thought that Hesperus is Phosphorus. In the first, a single concept is used twice. In the second, two concepts are each used once. This may have an impact on informativeness. When a concept is used twice in an identity thought, the thought is typically uninformative, whereas a thought similar in point of structure and content, but using two concepts, may be informative (Sainsbury and Tye 2011, p. 117).

(With respect to (FC), ST think difference in concepts explains rationality in *some* Frege cases, although not all of them. They would therefore deny that (FC) can be extended to all cases of contradictory belief. I will discuss and criticize their rejection of (FC) in sect. 4.2).

ST's account offers a clear example of an unrestricted individuation criterion for concepts, but several other RTM theorists have engaged in the same project. For instance, many authors subscribing to a non-Fregean ontology have tried to provide sets of *typing* conditions for concepts. This has been an especially popular programme among those working within the LoT framework: having identified concepts with LoT symbols, these theorists provide typing conditions for concepts by putting forward a typing criterion for LoT symbols. A typing criterion is a set of necessary and sufficient conditions under which two *token* concepts will belong to the same *type* concept. Here, for instance, is a summary of three possible typing criteria discussed by Aydede (2000a, sect. 1):

[Semantic]: two symbol tokens in different heads are of the same type if, and only if, they have the same semantic content.

[Physicalist]: two symbol tokens in different heads are of the same type if, and only if, they have the same physical [...] properties.

[Narrow functionalist]: two symbol tokens in different heads are of the same type if, and only if, they have the same narrow functional role¹⁹².

¹⁹² This can be equated with global inferential role.

And here is the “atomistic” typing criterion for concepts proposed by Fodor (1998), as it’s reconstructed by Aydede (1998, p. 289):

Fodor [...] distinguishes between concept identity and content identity, where content is understood to be broad (purely denotational). So different contents imply different concepts. But it is possible for two concepts to be type-distinct while identical in content (intra or interpersonally). The extra individuating element is what Fodor calls Modes of Presentation (MOPs), which are the vehicles that carry the content.

So, according to Fodor, the individuation condition for concepts is given by an ordered pair (a 2-tuple), whose first element is the broad content and the second a vehicle type that has the first as its semantic value: <denotation, vehicle type>. For present purposes we can represent concepts with these 2-tuples. According to Fodor, vehicles are terms in one's Language of Thought (LOT -- sometimes called Mentalese) realized in the brain. As such vehicles have both syntactic and semantic properties, and this fact can be used to answer the question raised by standard Frege cases: what makes co-denoting concepts type-distinct? (Aydede 1998, p. 289).

Fodor's view is well summarized by the formula:

[...] a mental representation is individuated by its form and content [...] (Fodor 1998, p. 27).

(Notice that Fodor and other authors in this tradition explicitly equate “typing conditions” and “individuation conditions”¹⁹³; I will follow their usage).

Other RTM-LoT theorists have proposed individuation criteria for LoT symbols that are of the same form as Fodor's but differ radically in content: v. for instance the holistic account recently proposed by Schneider:

[...] symbols must be individuated by their computational roles [...] (Schneider 2009a, p. 545).

[The Computational Theory of the Mind] requires a theory that types tokens by sameness and difference of total computational role, where the total computational role of a symbol is understood as the role it plays in the algorithms of a completed cognitive science (*ibid.*, p. 524).

Crucially, note how all these typing criteria amount to sets of sameness conditions that apply to *every* token concept: just like ST's originalism, they are “unrestricted individuation criteria” for concepts.

Interestingly, both atomists and holists want their criteria to predict that subjects involved in Frege cases will have distinct coreferential concepts for the relevant objects; in particular, they often take this to be required for a correct explanation of these subjects' behavior¹⁹⁴:

Morning Star beliefs have the same conditions of semantic evaluation as Evening Star beliefs, but they implicate the tokening of different syntactic objects and are therefore different beliefs with

¹⁹³ V. for instance Fodor (1998), chs. 1-2.

¹⁹⁴ V. especially Schneider (2009a, pp. 531-35), who is particularly explicit about this point; v. also Block (1993, 1995).

different causal powers [...] That believing P and believing Q may have different causal powers even if ‘P’ and ‘Q’ have the same semantic value shows up in all those operas where the soprano dies of mistaken identity (Fodor 1998, p. 39, my emphasis)¹⁹⁵.

(Of course, holists do *not* also take their criteria to be subject to the publicity constraint, while Fodor is probably the paradigmatic example of an RTM-LoT theorist who wants his account to make correct predictions about Frege cases while *also* complying with (PUB)¹⁹⁶).

The project of offering an individuation criterion for concepts has played an important role in the non-Fregean tradition. But authors on the Fregean side have also engaged in that project. Consider for instance the general form of the account defended by Peacocke (1992). According to Peacocke, a correct account of the possession (or “grasping”) conditions of a concept will also be a complete account of its individuation conditions:

Principle of Dependence: there can be nothing more to the nature of a concept than is determined by a correct account of the capacity of a thinker who has mastered the concept to have propositional attitudes to contents containing that concept (a correct account of “grasping the concept”) (Peacocke 1992, p. 5).

Given the Principle of Dependence, here is how an account of the individuation conditions for a particular concept will look like:

Accepting the Principle of Dependence opens up the possibility that we can simultaneously say in a single account what individuates a particular concept and also what it is to possess that concept. The general form that could be taken by such an account is this:

Simple Formulation: Concept F is that unique concept C to possess which a thinker must meet condition A(C) (*ibid.*, p. 6).

(For instance: the concept CONJUNCTION is that unique concept C to possess which a thinker must (roughly) be disposed to infer in conformity with the rules of conjunction introduction and elimination). Of course, a set of individuation conditions for a concept C will also be a set of *sameness conditions* for C: that is, any such set will amount to a set of necessary and sufficient conditions for a concept C* to be identical with C:

A statement that individuates a concept by giving its possession condition is an identity statement. The above statement of the possession condition for conjunction, for instance, is of the following logical form:

Conjunction = the unique concept C such that for a thinker to possess C is for ... C ... (*ibid.*, p. 9).

¹⁹⁵ V. also Fodor (1998, pp. 20-21).

¹⁹⁶ Aydede (1998; 2000 a,b) seems to share the same ambition.

Peacocke's account is thus an excellent example of an individuation criterion for concepts provided against the background of a Fregean ontology¹⁹⁷.

3.2. Some Terminology: Individuation, Restriction, Satisfaction

Having given a few concrete examples, we can finally offer a more general characterization of the notions of “unrestricted individuation criterion” and “constraint satisfaction”. I will call an “individuation criterion” for concepts any statement of a set of *necessary and sufficient sameness conditions* for concepts. A criterion will therefore include:

- A set of necessary conditions for two concepts C_1-C_2 to be the same concept.
- A set of sufficient conditions for two concepts C_1-C_2 to be the same concept.

(For instance: two concepts C_1-C_2 are the same concept iff they have same origin; two concepts C_1-C_2 are the same concept iff they have same reference and same syntactic properties; two concepts C_1-C_2 are the same concept iff they have same global inferential role; etc...). The conditions provided can of course take a number of different forms:

- The criterion might state a *single* necessary and sufficient condition for concept sameness (e.g. having same origin, according to ST).
- Alternatively, it might state a set of (two or more) *individually necessary and jointly sufficient* conditions. (On Fodor's account, for instance, having same reference *and* same syntactic properties are two individually necessary and jointly sufficient conditions for concept sameness).
- The conditions provided might also take *disjunctive* rather than *conjunctive* form: for instance, a criterion might state that two concepts C_1-C_2 are the same concept iff they have properties [F, G, H] *or* properties [I, J, K] *or* properties [F, G, K] *or* ...
- Crucially, the conditions provided might be *specific* to a particular concept rather than *general*. One might for instance provide the following necessary condition for a particular concept C_x ¹⁹⁸:

¹⁹⁷ Peacocke explicitly holds that a correct individuation criterion should account for puzzles of cognitive significance; v. for instance the already quoted “Distinctness of Concepts” principle (fn. 188 *supra*). He also subscribes to the idea that concepts are routinely shared by ordinary thinkers (Peacocke 1992, p. 3); this requirement is stressed even more by another Fregean, Georges Rey (v. Rey 1983, 1985, 1994, 2009a).

¹⁹⁸ A word about notation: I will generally use “ C_1 ” and “ C_2 ” as variables ranging over concepts in general, and “ C_x ” whenever I want to refer to a specific arbitrary concept.

A concept C is the same concept as $C_x \rightarrow$ concept C has property/properties [...]

(This is the kind of condition that will be provided on Peacocke's account. For instance, Peacocke holds that a concept C will be the same concept as CONJUNCTION just in case C has the possession conditions mentioned earlier. According to Peacocke, providing individuation conditions for *specific* concepts is one of the main tasks for a theory of concepts. His account is thus partially different from those previously examined, which all put forward *general* sets of sameness conditions¹⁹⁹).

A (necessary or sufficient) condition for concept sameness will be “unrestricted” just in case it applies to *all* concepts; an individuation criterion will be unrestricted just in case all of its sameness conditions are. The criteria examined in the previous section were all unrestricted: for instance, ST hold that, for *any* two concepts C_1-C_2 , they are the same concept iff they have same origin; the holist holds that, for *any* two concepts C_1-C_2 , they are the same concept iff they have same global inferential role; and so on. A restricted condition would be one in which the quantifier was restricted to a particular class of concepts, as in: for any two concepts C_1-C_2 *that are had by the same subject at the same time*, C_1-C_2 are the same concept iff they have same origin/global inferential role/etc.... Notice that this condition would be significantly *weaker* than the first. (We will say a lot more about restricted conditions in the next chapter. For now, I will simply take it for granted that all the conditions we'll examine are unrestricted).

Given the way I defined the relevant notions, an unrestricted individuation criterion (K) might consist of necessary and sufficient sameness conditions of the following form:

- (K1) For any two concepts C_1-C_2 : (C_2 is the same concept as $C_1 \rightarrow C_2$ has property/properties [...])
- (K2) For any two concepts C_1-C_2 : (C_2 has property/properties [...] $\rightarrow C_2$ is the same concept as C_1)

For instance, on ST's proposal:

- (K1) For any two concepts C_1-C_2 : (C_2 is the same concept as $C_1 \rightarrow C_2$ has **same origin** as C_1)
- (K2) For any two concepts C_1-C_2 : (C_2 has **same origin** as $C_1 \rightarrow C_2$ is the same concept as C_1)

In addition to general conditions of this kind, or in alternative to them, the criterion might include specific conditions for a particular concept C_x , of the form:

¹⁹⁹ V. Peacocke (1992, ch. 1) for more details.

- (K1_x) For any concept C: (C is the same concept as C_x → C has property/properties [...])
- (K2_x) For any concept C: (C has property/properties [...] → C is the same concept as C_x)²⁰⁰

(Cf. Peacocke's account. In this chapter I will mostly examine conditions of the first kind, i.e. the general conditions proposed by ST, Fodor, Schneider... ; later on, I will also consider conditions of the second kind).

Crucially, notice that since the expression “same concept” can be taken to express different kinds of “sameness relations”, an individuation criterion can be interpreted differently depending on what its proponent means by “same concept”. Suppose our theorist means “numerically identical concepts”: under this interpretation, his individuation criterion will state a set of necessary and sufficient conditions for two concepts to be numerically the same. (For instance, a Fregean theorist might provide a set of conditions under which two abstract concepts will be numerically identical). Alternatively, our theorist might mean “type-identical token concepts”: under this interpretation, his individuation criterion will state a set of necessary and sufficient conditions for two token concepts to belong to the same type, e.g. for two token LoT symbols to belong to the same LoT symbol-type. (This is how we should interpret the criteria proposed by non-Fregeans like Fodor, Schneider, Block etc...)²⁰¹. In what follows we will discuss individuation criteria of both kinds; whenever we examine a certain criterion, I will therefore make clear whether it should be interpreted in the numerical-identity or the type-identity sense.

We have now spelled out what an “unrestricted individuation criterion” for concepts amounts to²⁰². As anticipated, my main thesis is that, *if it satisfies (FC)*, a criterion of this form will:

²⁰⁰ Notice that the specific sameness conditions (K1_x)-(K2_x) are *also* unrestricted, just like the general conditions (K1)-(K2). The distinction between general/specific conditions should not be confused with the distinction between unrestricted/restricted conditions.

²⁰¹ It's not entirely clear whether ST intend to provide a set of identity conditions or just a set of typing conditions; fortunately, we need not decide the issue here, since my objections against their account go through on both readings.

²⁰² We might also add some further requirements: for instance, we might want our criterion to be “minimal”. If it's true that “C₂ is the same as C₁ only if C₂ has property/properties [...]”, then it's also true that “C₂ is the same as C₁ only if C₂ has property/properties [...] *and* 2 + 2 = 4”. By requiring the account to be minimal, such intuitively irrelevant conditions would be ruled out. We might also require the account to be “informative” and maybe include some further constraints as well. I am unable to spell out in detail what these requirements could amount to; I hope the examples considered in section 3.1, together with the general characterization offered in this section, will be enough to give the reader an intuitive sense of the kind of sameness conditions we are interested in.

- Either violate (PUB);
- Or at any rate entail that several concepts (those involved in Frege cases) are not shared.

But what do I mean by “satisfying (FC)?” Before offering a more precise characterization of the notion, let me give you a more intuitive one. A sameness condition (and the individuation criterion embedding it) will satisfy one of our constraints just in case it makes *the right predictions* about concept sameness and difference in all the cases covered by that constraint. That is: whenever the publicity constraint entails that, in a certain specific case, two concepts C_X-C_Y are the *same* concept, it should follow from our individuation criterion that C_X-C_Y are indeed the same; and whenever the Fregean constraint entails that, in a certain specific case, two concepts C_X-C_Y are *different* concepts, it should follow from our individuation criterion that C_X-C_Y are indeed different. This means that, in order for an individuation criterion to satisfy a constraint, it will have to include one or more sameness conditions that make the right predictions about the relevant cases (so I’ll talk about sameness conditions and individuation criteria “satisfying” a constraint more or less interchangeably).

Let’s give an example. (FC) entails that the concepts involved in Lois’ contradictory beliefs about Superman/Clark are different; as usual, we will call these two concepts SUPERMAN and CLARK. Since it follows from (FC) that these two concepts are distinct, an individuation criterion (K) will satisfy (FC) *only if* it follows from (K) that Lois’ SUPERMAN is a different concept from Lois’ CLARK. This means that, in order for (K) to satisfy (FC), it must include a *necessary* condition for concept sameness which is *not* met by Lois’ two concepts, in which case it will indeed follow from (K) that the two concepts are distinct. For instance: if (K) included the necessary condition “ $C_1 = C_2$ only if they have same global inferential role”, then it would clearly follow from (K) that Lois’ concepts are distinct. Similarly for all other “Frege cases” covered by (FC): an individuation criterion (K) will satisfy (FC) iff, for *each* of these cases, (K) entails that the concepts involved in the subject’s contradictory beliefs are different²⁰³. (The same holds for publicity, although my focus in the next chapters will be on (FC))²⁰⁴.

We can now offer a more formal definition of the notion of “satisfaction” (I will only define the notion with respect to (FC)):

²⁰³ Notice that (FC) can also be satisfied through a set of concept-specific conditions rather than a general one. For instance, to satisfy (FC) in Lois’ case, our criterion could include the specific condition “If a concept $C = \text{SUPERMAN}$, then C is inferentially connected to [HAS SUPERPOWERS]”: since this condition is not met by Lois’ CLARK, our criterion would satisfy (FC) in this case. Further specific conditions would then have to be given for every other Frege case.

²⁰⁴ Notice that “satisfying” a constraint is *not the same* as being consistent with it; a sameness condition can be consistent with a constraint but fail to satisfy it, i.e. fail to make the right predictions about the cases covered by that constraint. We will consider several instances of this later.

Satisfying (FC): A necessary condition for concept sameness (K1) *satisfies* (FC) iff, for any world w in which two concepts C_X-C_Y are involved in the contradictory beliefs of a rational subject S at t about some x , it follows from (K1) *plus* a complete description of w that C_X is a *different concept* from C_Y .

(An individuation criterion will then satisfy (FC) just in case it includes a necessary condition, or a set of necessary conditions, which satisfies (FC)). Consider a world w in which two concepts C_X-C_Y are involved in the contradictory beliefs of a rational subject S at time t about some object x (e.g. the world in which SUPERMAN and CLARK are involved in Lois' contradictory beliefs about Superman/Clark). If a necessary condition *doesn't* entail that C_X is different from C_Y (given a complete description of w), then it doesn't satisfy (FC): we now have a case in which it follows from (FC) that two concepts are distinct, but in which (K1) doesn't entail that they are. If on the other hand our necessary condition *does* entail that C_X is different from C_Y (given a complete description of w), then it *does* satisfy (FC) in w ; and if it also satisfies (FC) in all the other worlds in which (FC) applies, then it just satisfies (FC) in general. Going back to our usual example, the necessary condition: “ $C_1 = C_2$ only if they have same global inferential role” clearly satisfies (FC) in Lois' case: w is a world in which Lois' two concepts do *not* have same global inferential role, so it follows from our condition + w 's description that Lois' SUPERMAN is different from Lois' CLARK, which is just what (FC) requires.

4. A Tension Between Constraints

4.1. The Double Superman Case

In this section, I will consider a case in which:

- (FC) requires two concepts C_X - C_Y to be distinct;
- But all the unrestricted necessary conditions which correctly predict that C_X is distinct from C_Y also violate (PUB)²⁰⁵.

The case raises a puzzle: can we find an unrestricted criterion that satisfies (FC) while being *consistent* with publicity? At the end of this section, I will consider a first possible solution to the puzzle and note how that solution *still* entails that many concepts are not shared; whether this is consistent with publicity is something I will discuss in the next chapter.

The case I'll discuss is a simple variant of our standard Frege case involving Lois and Superman/Clark. As we'll see, the case is not at all far-fetched or extraordinary: analogous Frege cases arise all the time in our everyday cognitive lives. In Lois' case, a rational subject has two concepts that are involved in contradictory beliefs *and* are respectively associated with the "standard" Superman-ish role and the "standard" Clark-ish role: Lois believes of Superman/Clark that he can fly under the first concept, and that he cannot fly under the second. What makes the inferential roles "standard", of course, is the fact that most inhabitants of Metropolis possess concepts having the roles in question (the superhero called "Superman" and the reporter called "Clark Kent" are both well-known characters).

Things are a bit different in our case: here, a rational subject has two concepts that are involved in contradictory beliefs about Superman/Clark and are *both* associated with the standard Superman-ish role. Each concept refers to Superman/Clark and is inferentially connected to [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY, HAS SUPERPOWERS...]; just like Lois, however, our subject rationally ascribes to Superman/Clark the property of being F under the first concept, and the contradictory property of being not-F under the second.

It's easy to imagine a case like this. Suppose Edmund is a rational, normal subject under all relevant respects. Like Lois and most other inhabitants of Metropolis, he believes there is a superhero called "Superman" who wears a red cape, can fly, has superpowers and so on. At some point, Edmund hears on the news that, within a few minutes, someone fitting his Superman-description captured a bank robber and saved a kitten in two very distant parts of Metropolis. As a result, he concludes (reasonably, but mistakenly) that there must be two *distinct* people who are both called "Superman",

²⁰⁵ By "violate" I simply mean "are inconsistent with".

wear a red cape, can fly etc... : one of them captured the bank robber, the other saved the kitten. So Edmund believes, of the person who appeared in zone A of Metropolis, that he captured the robber; at the same time, he believes of the person who appeared in zone B that he did *not* capture the robber. Therefore, Edmund believes of Superman that he both did and did not capture the robber.

Clearly, Edmund is not irrational in his contradictory *de re* beliefs. If so, then (FC) entails that the concepts involved in those beliefs must be *distinct concepts* (if they weren't, we could not apply the Standard Explanation to account for Edmund's rationality). Let's call the two concepts SUPERMAN₁/SUPERMAN₂. Given our description of the case, it follows that:

- Both concepts have the standard Superman-ish role [CALLED “SUPERMAN”, WEARS A RED CAPE, CAN FLY...].
- However, SUPERMAN₁ is also connected with [CAPTURED THE BANK ROBBER, DID NOT SAVE THE KITTEN].
- While SUPERMAN₂ is connected with [DID NOT CAPTURE THE BANK ROBBER, SAVED THE KITTEN].

(You can also think of the two concepts as two “mental files”: each file “contains” the standard Superman-description, but it also contains some information about Superman which contradicts the information contained in the other file²⁰⁶). Since (FC) entails that SUPERMAN₁ is different from SUPERMAN₂, an individuation criterion that satisfies (FC) must include a necessary condition for concept sameness which, together with a complete description of the case, entails that SUPERMAN₁ is indeed different from SUPERMAN₂. We must therefore look for an unrestricted necessary condition which is *not* met by Edmund's two concepts: this condition will entail that SUPERMAN₁ is different from SUPERMAN₂, and (FC) will be satisfied. While looking for this condition, our driving question will be: can we find an unrestricted condition which entails that SUPERMAN₁ is different from SUPERMAN₂ *and is at the same time consistent with (PUB)*? As we will see in the next section, all the major individuation criteria examined so far seem to fail on at least one of these scores.

4.2. Satisfying (FC) Vs. Respecting (PUB)

In this section, I will consider four unrestricted conditions for concept sameness, and note that each of them will *either* fail to satisfy (FC) in Edmund's case *or* violate publicity:

²⁰⁶ Cf. Recanati (1993, 2009) and Perry (2000) for an overview of the mental files framework.

- a) *For any two concepts C_1 - C_2 : (C_2 is the same concept as $C_1 \rightarrow C_2$ has same reference as C_1)*

Clearly, (a) is a plausible condition on concept sameness: many RTM theorists hold that concepts are at least partially individuated by their reference. While we might well want to include (a) in our individuation criterion, however, this will clearly not be enough to satisfy (FC) in Edmund's case (or indeed in any Frege case): SUPERMAN₁-SUPERMAN₂ have same reference, so (a) fails to entail that they are distinct. We must therefore look for some *further* necessary condition which entails that SUPERMAN₁ is distinct from SUPERMAN₂ without flouting (PUB). This is what will prove hard; to see why, let's turn to the individuation criteria examined in section 3.1.

- b) *C_2 is the same concept as $C_1 \rightarrow C_2$ has same origin as C_1* ²⁰⁷

To determine whether ST's account can deal with our case, we must say a bit more about the notion of an "origin". According to ST, concepts "come into existence": they are "non eternal continuants" (2011, p. 102). A concept can be created in at least two ways. First, someone might introduce a new term in public language to refer to some object x : he thereby creates a concept C which refers to x and can be acquired by other subjects when they learn the term associated with C . Acquiring a concept in this way will involve deferring to previous users: "deference takes the form of intending to use the concept as it has been used by oneself or others on previous occasions" (*ibid.*, p. 103). A thinker might also create a concept without associating it with a public term. This "private" concept might be introduced through an "explicit intentional introduction", but in many cases it will simply be the product of subpersonal processes: this will be the case for most of the concepts infants acquire in the course of their development. Crucially, however, ST think that "the concepts infants form on their own are typically supplanted by public concepts when they become full members of their surrounding linguistic community [...] as the child becomes a member of his conceptual community, [his private concept for cats] will be supplanted by one or more public concepts, for example the concept CAT [...]" (*ibid.*, p. 104).

As it stands, ST's originalist theory seems to flagrantly violate the publicity constraint; moreover, the account doesn't even satisfy (FC) in all the relevant Frege cases. Let me address each of these points in turn. The problem with publicity is easily seen: if sameness of origin is necessary for concept identity, then whenever two concepts are associated with terms introduced by different linguistic communities they will be different concepts. If my current concept CAT is the one I acquired through linguistic immersion, by deferring to uses of previous speakers who employed the English term "cat", then it is *not the same concept* as the one associated with the

²⁰⁷ Throughout the rest of this section, I'll leave the quantifier "for any two concepts C_1 - C_2 " implicit when stating sameness conditions.

Spanish word “gato”. So, assuming I don’t speak Spanish and have never come across the word “gato”, I don’t have the concept GATO. Therefore, a Spanish speaker who is in the same position with respect to the English word “cat” will not have any concept for cats in common with me. Notice, however, that me and my Spanish counterpart are certainly covered by many of the same cat-related generalizations, and that we genuinely agree about many features of cats. Therefore, (PUB) entails that we do have a concept for cats in common: ST’s necessary condition for concept sameness is too strict and makes wrong predictions about speakers belonging to different linguistic communities. This is ironic, since the main motivation for claiming that public concepts supplant the infant’s private concepts is to allow for concept publicity within a certain linguistic community²⁰⁸: unfortunately, (PUB) also requires concepts to be shared across different communities.

Not only is ST’s necessary condition for concept sameness too fine-grained for publicity purposes; it also seems too coarse-grained for Fregean purposes. There are cases in which (FC) requires certain concepts to be distinct, even though they have the *same origin*: in all these cases, ST’s account will fail to make the right predictions. Admittedly, originalism will satisfy (FC) in some Frege cases. For instance: if James believes that Cicero was a Roman orator and that Tully was not a Roman orator, his beliefs will presumably involve two distinct concepts CICERO/TULLY, which are associated with different public names and which therefore have distinct origins²⁰⁹; similarly for Hesperus/Phosphorus cases and other analogous ones. In *some* scenarios, however, the concepts involved in a subject’s contradictory beliefs will *not* have different origins. For instance, ST hold that, in Kripke’s Paderewski case, Peter accepts the thought PADEREWSKI HAS MUSICAL TALENT and *also* accepts the negation of the very *same* thought, i.e. PADEREWSKI DOES NOT HAVE MUSICAL TALENT (v. Sainsbury and Tye 2012, pp.131-138). When Peter becomes acquainted with Paderewski at the concert, he learns the proper name “Paderewski”, acquires the public concept PADEREWSKI and comes to accept a thought involving that concept, i.e. PADEREWSKI HAS MUSICAL TALENT. Later, at the political rally, Peter encounters the same name “Paderewski” and comes to accept a thought involving the *same* public concept, i.e. PADEREWSKI DOES NOT HAVE MUSICAL TALENT. Edmund’s case is analogous. When Edmund hears the news about Superman’s deeds in zone A and then about his deeds in zone B, one and the same public name (“Superman”) is used twice by the news reporter. So, on ST’s account, Edmund will come to accept a thought (SUPERMAN CAPTURED THE ROBBER) and the negation of the very same thought (SUPERMAN DID NOT CAPTURE THE ROBBER).

In cases like Peter and Edmund’s, (FC) entails that different concepts are involved in the subject’s thoughts, while ST hold that one and the same concept is involved. Now, ST do *not* take this to be a problem for their account, since they would *reject*

²⁰⁸ As noted earlier, ST explicitly accept standard RTM arguments in support of publicity: v. the passage quoted in sect. 3.1 *supra*. V. also Sainsbury and Tye (2012, p. 87-88), where they argue that their view is superior to Millikan’s in its allowing for shared concepts.

²⁰⁹ Sainsbury and Tye (2012, pp. 12-13).

(FC). In some cases involving contradictory *de re* beliefs, a subject's rationality is indeed explained by his accepting thoughts involving distinct concepts; v. for instance the Cicero/Tully and Hesperus/Phosphorus cases described above. In cases such as Peter or Edmund's, however, a subject is rational even though his contradictory beliefs involve the *same* concept: these cases show (FC) to be false.

How, then, will ST explain Peter/Edmund's rationality, if not by appealing to the distinctness of their concepts? ST's strategy is to appeal to their beliefs *about* their own concepts: "Peter thinks, at the rally, that he has learned a new concept PADEREWSKI, different from the PADEREWSKI concept he had earlier acquired at the concert. He is wrong. But if this is a reasonable mistake, it is one that makes it reasonable for him to believe that his thoughts are not contradictory. *He reasonably believes that it is not the case that the one thought consists in the other embedded in a concept for negation*" (*ibid.*, p. 134, my emphasis).

I cannot make full justice to ST's argument here. My main purpose in this section is to show that certain individuation criteria will violate (PUB), even when they *do* satisfy (FC); and I have already shown ST's account to be inconsistent with publicity. Let me note, however, that there are at least two problems with ST's strategy. First, it seems strange to offer *different* accounts of rationality in some Frege cases but not others, depending on how many public names are involved. Why would James' rationality be explained by the fact that his contradictory beliefs involve distinct concepts CICERO/TULLY, while Peter's rationality is explained by his meta-conceptual beliefs about his PADEREWSKI-concept? For this asymmetry to be justified, the fact that a single public name "Paderewski" is involved in Peter's case should somehow make a difference. But why would the fact that Paderewski only happened to have one name, while Cicero had two, determine a difference in what makes Peter and James rational? To put things differently: if what explains James' rationality is the fact that his beliefs involve distinct concepts, it seems that the same fact should also explain Peter's rationality. At the very least, the resulting account would have the advantage of being more unified than ST's.

An even more serious worry is that, *contra* ST, Peter simply doesn't seem to have *any* beliefs about the identity of his concept PADEREWSKI. *Prima facie*, Peter does not believe that he acquired a concept at the concert and a second, distinct concept at the rally, nor that these concepts are involved in the thoughts he formed on those occasions. Peter certainly never makes claims about the identity of his own concepts, and we might suppose he is an entirely unsophisticated thinker. More generally, it seems implausible and *ad hoc* to claim that *all* subjects who find themselves in a Paderewski-like case must have meta-conceptual beliefs like those invoked by ST²¹⁰.

In reply, ST could try to weaken their claim: perhaps all that's needed to explain Peter's rationality is that he *doesn't believe* his two thoughts involve the same concept, *not* that he positively believes they don't. But this seems too weak. Suppose Peter learns

²¹⁰ Interestingly, ST seem to agree: v. Sainsbury and Tye (2012, p. 136).

of the identity between the pianist and the politician, thus coming to accept claims like “The man called ‘Paderewski’ who was at the concert is the same person as the man called ‘Paderewski’ who was at the rally”. Suppose, however, that Peter also retains his contradictory beliefs after becoming enlightened: for instance, suppose he keeps saying things like “The man called ‘Paderewski’ who was at the concert has musical talent” and “The man called ‘Paderewski’ who was at the rally lacks musical talent”. Clearly, Peter would be *irrational* in these circumstances. But, if Peter didn’t believe that his inconsistent thoughts involved the same concept *before* becoming enlightened (e.g. because he is an unsophisticated thinker), there is no reason why he should have acquired that belief *now*: being unsophisticated, Peter will lack meta-conceptual beliefs at *any* time. So the weakened version of ST’s account incorrectly predicts that Peter is *still* rational after he learns the identity: after all, he still doesn’t believe that his thoughts involve the same concept PADEREWSKI.

I conclude that ST’s originalist condition on concept sameness violates (PUB); that it fails to satisfy (FC) in some Frege cases; and that ST’s rejection of (FC) forces them to adopt a problematic account of rationality in those cases.

c) C_2 is the same concept as $C_1 \rightarrow C_2$ has same inferential role as C_1

This is the sameness condition endorsed by Inferential Role Semantics (IRS) theories of concepts. To determine what kind of predictions it will make in our case, we must of course give a definition of “inferential role” first. As we have seen in chapter 2²¹¹, there are two options here, and therefore two possible ways to decline (c). First, we considered a *holistic* definition of “inferential role”, on which the role of a concept C is the set of *all* the inferential connections in which C stands at a time t for a subject S . On a holistic construal, (c) clearly satisfies (FC) in Edmund’s case, since his SUPERMAN₁-SUPERMAN₂ differ in many of their inferential connections. However, it’s equally clear that (c) will then violate (PUB), for the usual reasons: on a holistic definition of “inferential role”, two subjects won’t have any concepts in common unless they share all of their beliefs.

A second possibility would be to give a non-holistic, or “*localist*” construal of inferential roles²¹²:

Localist definition of inferential roles: the inferential role of a concept C is the set of *some* (but not *all*) the inferential connections in which C stands at a time t for a subject S .

²¹¹ Sect. 2.3.

²¹² Localist theories of concepts are defended by (among others): Weiskopf (2009 a,b), Devitt (2006), Pereboom (1995), Rey (1983, 1985, 2005), Peacocke (1992, 1998a, 2007). The formulation in the text simplifies some of the differences between localist views, but not in a way that matters here. In chapter 6 I will discuss different varieties of localism in more detail.

Of course, our necessary condition (c) will now be a lot less strict than on the holist's reading. Suppose the local inferential role of BACHELOR is [UNMARRIED MAN]. The localist will then hold that, even if I happen to believe that all bachelors are lonely and you don't, we can still share the same concept BACHELOR. As long as we are both disposed to infer from BACHELOR to UNMARRIED MAN (and vice versa), our two concepts will be the same even if they differ in some of their connections (e.g. BACHELOR → LONELY). For this reason, localist versions of IRS seem better equipped to deal with publicity (arguably, one of the main motivations for the view). But can they also satisfy (FC)?

Before discussing Edmund's case, notice that localism *does* seem to satisfy (FC) in more "standard" Frege cases like Lois'. For instance, suppose we identify the inferential role of the concept SUPERMAN with the standard Superman-ish role: on this view, a concept C will be the concept SUPERMAN only if it stands in an inferential connection to [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY...]. This would enable the localist to satisfy (FC): Lois has a concept C_1 for Superman/Clark which is associated with [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY...] and a concept C_2 which is not (it is associated with [CALLED "CLARK", SHY, WEARS GLASSES...]). It thus follows from localism that Lois' C_2 is not the same concept as her C_1 , since the two concepts have different local roles.

At the same time, a localist necessary condition does not seem to violate (PUB) in Lois' case. Given that most people associate the *same* description with the name "Superman", they all have a concept C that has the inferential role of SUPERMAN; therefore, they all meet the localist necessary condition on concept sameness. Whenever publicity requires two subjects to share SUPERMAN for purposes of communication, agreement and so on, the localist condition will be respected. Localism thus seems to give us the best of both worlds: it cuts concepts finely enough to satisfy (FC), and without violating (PUB) at the same time!

Now, I believe this is a mistaken diagnosis determined by some exceptional features of Lois' case, most notably the fact that the inferential role of Lois' two concepts happens to be widely shared by many other inhabitants of Metropolis. As Edmund's case shows, however, other Frege cases will be very different. By hypothesis, Edmund's SUPERMAN₁-SUPERMAN₂ are both inferentially connected with [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY...]. Unlike Lois' SUPERMAN-CLARK, then, Edmund's concepts *share* the standard Superman-ish role: they only differ in certain other, "non-standard" inferential connections ([CAPTURED THE ROBBER/DID NOT CAPTURE THE ROBBER]; [DID NOT SAVE THE KITTEN/SAVED THE KITTEN]²¹³). Therefore, Edmund's concepts do *not* seem to violate our necessary condition (c). If we define "inferential role" as the localist proposes, the two concepts seem to have the *same inferential role*,

²¹³ Of course, Edmund will make various inferences from these features, which will result in further differences between the inferential connections of the two concepts.

i.e. the inferential role of the (widely shared) concept SUPERMAN. If so, our localist condition (c) will not satisfy (FC) in Edmund's case.

A first possible localist reply would be that we have mischaracterized the inferential role of Edmund's concepts; that role is a lot more specific than the standard Superman-ish role. For instance, we could describe the role of the two concepts as follows:

- i. **Inferential role of SUPERMAN₁:** [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY, HAS SUPERPOWERS..., CAPTURED THE ROBBER, DID NOT SAVE THE KITTEN]
- ii. **Inferential role of SUPERMAN₂:** [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY, HAS SUPERPOWERS..., DID NOT CAPTURE THE ROBBER, SAVED THE KITTEN]

Clearly, the inferential role of Edmund's SUPERMAN₁ is now *different* from the inferential role of his SUPERMAN₂. It will then immediately follow from our necessary condition (c) that SUPERMAN₁ is not the same concept as SUPERMAN₂, since they have different inferential roles. And, crucially, this will be compatible with a localist definition of "inferential role", since the roles in (i)-(ii) need *not* include *all* the inferential connections in which Edmund's concepts stand. So the inferential role of the two concepts is not holistic. Conclusion: (c) *does* satisfy (FC), even assuming a localist reading of the condition.

The problem with the localist's response is easy to see. To distinguish between SUPERMAN₁ and SUPERMAN₂, the localist must include very specific information (such as [CAPTURED THE ROBBER]) in their inferential roles. Once we do that, however, we will run into trouble with publicity again: if the inferential role of my concepts is *that* specific, then there will almost certainly be a subject covered by (PUB) whose concepts *do not* have the same inferential role as mine. To see why, let's elaborate on our case a bit more. Suppose the inferential role of Edmund's two concepts is the one described above, and consider Verna, who hasn't heard the news about the last deeds of the superhero called "Superman" (she doesn't know Superman captured the robber, she doesn't know Superman saved the kitten, etc...). Verna does have one or more concepts of Superman with the standard inferential role [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY...]. Because she hasn't heard the news, however, none of these concepts is inferentially connected to [DID NOT CAPTURE THE ROBBER, SAVED THE KITTEN] or to [SAVED THE KITTEN, DID NOT CAPTURE THE ROBBER]. So (c) entails that *none* of Verna's concepts for Superman is the same as any of Edmund's concepts. But, surely, (PUB) *does* require Verna and Edmund to have a concept for Superman in common; after all, Verna and Edmund can agree/disagree about Superman, successfully exchange information about him, be covered by the same Superman-related generalizations, and so on. Conclusion: the localist response might enable (c) to satisfy (FC), but it also makes (c) incompatible with (PUB).

Before we consider a further reply that's available to the localist (sect. 4.3 *infra*), I would like to discuss one last candidate.

d) *C₂ is the same concept as C₁ → C₂ belongs to the same symbol-type in the Language of Thought (LoT) as C₁*

This is the necessary condition for concept sameness proposed by non-Fregean ontologists such as Fodor and Schneider. On their version of RTM, SUPERMAN₁-SUPERMAN₂ are two token symbols in Edmund's LoT. This claim can be roughly fleshed out as follows: Edmund's two concepts are token mental representations realized by neurological states of his brain; in turn, these states have a set of physical-neurological properties to which thought processes involving the two concepts are sensitive.

If Edmund's two concepts belong to different LoT symbol-types, it will follow from (d) that they are not the same concept. But what are the typing criteria for LoT symbols? This is a controversial issue: as we have seen, different options are available²¹⁴. A first possibility would be to type LoT tokens by their *global computational role*, i.e. by their global set of inferential connections with other concepts, as well as perceptual inputs and behavioral outputs. This is the “holistic” typing criterion suggested by Schneider: needless to say, this criterion would entail that different subjects almost never have token concepts of the same type, which would constitute a straightforward violation of (PUB)²¹⁵.

A second possibility would be to type LoT tokens by their reference *and* their *syntactic properties*, as suggested by Fodor (1998, 2008). But this creates a new dilemma: what are the syntactic properties of a token LoT symbol? We seem to have two main options: we can either identify the syntactic properties of a symbol with its *physical-neurological* properties, or with its *computational* properties. Unfortunately, no matter which option we choose, Fodor's account will raise just the same problems as the candidates we already discussed²¹⁶.

Suppose we identify the syntactic properties of a LoT token with its physical-neurological properties. This would not be unmotivated: one of the central tenets of the LoT hypothesis is that computational processes are defined over the *syntactic properties* of the symbols on which they operate. But if computations must be implemented by a *physical* mechanism (be it a Turing machine or a brain), the operations of that mechanism can only be sensitive to the *physical properties* of the symbols on which the mechanism operates. In particular, the mechanism will operate differently on two token symbols A-B only if they differ in some of their physical properties; if they didn't, there would be no way for the mechanism to “read” their difference and treat them differently in its computations. Once we conjoin these central tenets of the LoT hypothesis, it follows that we should identify the syntactic properties of a token symbol with its

²¹⁴ V. Schneider (2009 a,b, 2011) for discussion.

²¹⁵ V. ch. 3 (sect. 4.2) for discussion of Schneider's account.

²¹⁶ V. Aydede (1998, 2000 a,b), who offers a detailed argument against Fodor along the same lines. My critique of a “neurological” account of syntactic properties is based on Aydede's argument; more generally, many of the ideas in chs. 2-5 were inspired by Aydede's careful treatment of these topics. V. also Millikan (2011, pp. 128-129).

physical properties. This seems to yield the right predictions about our case, as well: presumably, Edmund is in different brain states when thinking thoughts involving SUPERMAN₁-SUPERMAN₂, so this version of Fodor's proposal does satisfy (FC). Unfortunately, however, it also violates (PUB). When thinking about some object x , different thinkers will almost invariably be in brain states that differ in some of their physical-neurological properties (not to mention the fact that this proposal would block the possibility of sharing concepts with young infants, silicon computers, aliens and so on²¹⁷). And even if we could show that some intentional subjects are indeed in the same brain state when thinking certain thoughts, it would be clearly absurd to maintain that two thinkers will be in exactly the same brain state *whenever* they are covered by the same generalizations, communicate successfully, or agree/disagree on something.

One might then be tempted to move to a “functionalist” account and hold that the syntactic properties of a token symbol must be identified with its computational properties, i.e. with the property it has of standing in certain computational connections with other concepts, perceptual inputs and behavioral outputs. It will then follow from Fodor's account that two token concepts belong to the same LoT symbol-type only if they have the same computational role²¹⁸. Clearly, however, our necessary condition (d) will then collapse into the previously discussed condition (c), since the computational role of a concept *just is* its inferential role²¹⁹. The same dilemma will then arise. On a holistic definition of “computational/inferential role”, (d) will violate (PUB) (notice that Fodor's typing criterion would collapse into Schneider's holism on this version of the account); on a localist definition, it will simply fail to satisfy (FC), since it will not predict that Edmund's concepts are distinct. Ironically enough, Fodor's account seems to face exactly the same problems as IRS theories once we try to get clearer on what the “syntactic properties” of a concept amount to.

4.3. A Solution?

I have considered four plausible necessary conditions for concept sameness and shown that they *either* violate publicity *or* fail to satisfy the Fregean constraint in Edmund's case: it seems impossible to find an individuation criterion for concepts that satisfies (FC) while being consistent with (PUB).

Now, there is at least one possible solution I haven't discussed. To get an idea of how the solution might go, consider the following “modified localist” proposal. The localist, recall, accepts the IRS condition (c) on concept sameness (“If C_2 is the same concept as $C_1 \rightarrow C_2$ has same inferential role as C_1 ”) and combines it with a localist definition of “inferential role”. To satisfy (FC), our localist might hold that the local

²¹⁷ Fodor 2008, p. 90.

²¹⁸ Schneider (2009 a,b) proposes various arguments in defense of a functionalist account of LoT syntax.

²¹⁹ In fact, computational roles are even more specific than proper inferential roles, since they also include computational connections to perceptual inputs/behavioral outputs.

inferential role of Edmund's SUPERMAN₁-SUPERMAN₂ is the *very specific* one described previously:

- i. **Inferential role of SUPERMAN₁:** [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY, HAS SUPERPOWERS..., CAPTURED THE ROBBER, DID NOT SAVE THE KITTEN]
- ii. **Inferential role of SUPERMAN₂:** [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY, HAS SUPERPOWERS..., DID NOT CAPTURE THE ROBBER, SAVED THE KITTEN]

In response to my objection, however, the localist might hold that not *all* concepts have equally specific roles. For instance, there might be a *third* concept (which we can simply call "SUPERMAN") whose inferential role is much less specific than that of Edmund's SUPERMAN₁-SUPERMAN₂. In particular, we might suppose that SUPERMAN's inferential role only includes the standard Superman-ish features [CALLED "SUPERMAN", WEARS A RED CAPE, CAN FLY...]. By hypothesis, Verna *does* have such a concept, since she does believe, of the guy called "Superman", that he satisfies the standard Superman-description. So Verna does have a concept that has same inferential role as one of Edmund's concepts. Therefore, the two subjects can share the concept SUPERMAN and there is no violation of publicity. At the same time, however, our localist condition (c) will make the right predictions about SUPERMAN₁-SUPERMAN₂ being different, since the two concepts have the very specific inferential roles described above. In a nutshell, a localist might hold that, while concepts are indeed individuated by their local inferential roles, there is no reason to think that all roles will have the same degree of specificity; some of them (SUPERMAN) will be more coarse-grained, while others (SUPERMAN₁-SUPERMAN₂) will be more fine-grained²²⁰.

As you might have noticed, this solution can in fact be applied not only to localism, but to pretty much *all* the individuation criteria examined in this section. For instance, a holist could adopt the following position: Edmund's concepts SUPERMAN₁-SUPERMAN₂ are holistically individuated and not shared, so as to satisfy (FC); at the same time, however, the concept SUPERMAN is *not* holistically individuated and can be shared by

²²⁰ As usual, the move will have to be declined differently depending on our background ontology for concepts. A *Fregean* can simply hold that Edmund has two *abstract* concepts SUPERMAN₁-SUPERMAN₂ with the inferential roles described above; at the same time, he also has a third abstract concept SUPERMAN, whose inferential role is the standard Superman-ish one and which is therefore shared between him and Verna. A *non-Fregean* could instead hold that there are two distinct *type* concepts SUPERMAN₁-SUPERMAN₂, where the inferential role of SUPERMAN₁ includes [... CAPTURED THE ROBBER, DID NOT SAVE THE KITTEN] and the inferential role of SUPERMAN₂ includes [... DID NOT CAPTURE THE ROBBER, SAVED THE KITTEN]. A token concept will then belong to one of these types just in case it has the inferential role individuating that type; since only one of Edmund's token concepts for Superman has the former inferential role, and only one of them has the latter, the two concepts belong to distinct types and (FC) is satisfied. In addition to SUPERMAN₁-SUPERMAN₂, however, there is also a third type concept SUPERMAN: this concept's inferential role is just the standard Superman-ish one, and both Verna and Edmund have a token concept with that role.

Edmund and Verna, so that publicity is respected²²¹. *Mutatis mutandis* for other individuation criteria: all we will have to do is combine our favored individuation method with the idea that concepts are individuated in different ways, some ($SUPERMAN_1$ - $SUPERMAN_2$) more finely and some ($SUPERMAN$) more coarsely.

In the next chapter, we will say a lot more about how to apply this solution to an account of concept individuation/possession. For now, let me just note that the solution still has the consequence that several concepts are *not shared*: on the strategy I just sketched, Edmund's $SUPERMAN$ will be shared by Verna, but his $SUPERMAN_1$ - $SUPERMAN_2$ will not. More generally, many concepts involved in contradictory beliefs will be individuated so finely that several intentional subjects will *not* have them, *even though* these subjects communicate, agree etc... with the owner of the concepts in question. So our solution still entails that there are several concepts which certain intentional subjects do not share, even though these subjects are all covered by (PUB). Whether this would be *compatible* with the publicity principle is a question I will go back to in the next chapter.

²²¹ Notice that this is the “weakened” holistic view described in ch. 3 (sect. 4.1): we will go back to this in the next chapter.

Chapter 5

Solving The Tension

1. Introduction

Let's take stock. In the previous chapter I introduced a new constraint on a theory of concepts and showed that, once we accept this constraint, a problem will arise: unrestricted individuation criteria for concepts that satisfy the constraint also appear to violate publicity. I examined several existing criteria for concept individuation and argued that they all have to face this worry. Then, I considered a possible way to "patch" some of those criteria (sect. 4.3). I remarked that, even when patched, our criteria would *still* entail that several concepts are not ordinarily shared, i.e. those concepts that are involved in the contradictory beliefs of subjects in Frege cases; I left open, however, whether this consequence would be compatible with the principle of publicity.

I will soon go back to the solution sketched at the end of chapter 4, comparing it with other possible strategies and trying to determine whether it would be consistent with (PUB). Before we discuss how to solve our problem, however, there are two points that need to be addressed:

- First (sect. 2), I want to analyze in more detail the *implications* of (FC). I will show in a principled way that, once we accept (FC), we will be *forced* to adopt a very fine-grained sameness conditions for concepts. This is something we will have to take into account when looking for an individuation criterion that can respect our two constraints.
- Second (sect. 3), I will try to identify the *causes* of the problem described in chapter 4. I will show why an unrestricted individuation criterion is very likely to either violate publicity (as the candidates in sect. 4.2) or at any rate entail that the concepts involved in Frege cases are not shared (as the "patched" candidates in sect. 4.3). This will give us principled reasons to think that *any* unrestricted criterion would have one of these consequences, not just those examined in chapter 4.

After our discussion in sections 2-3, we will be in a better position to find a solution to our problem. In section 4, I will examine two possible ways to solve the tension between publicity and (FC). My main goal will not be to determine which one is better (although I will express a preference for one of the two). I will rather try to provide support for my two main theses; as you might recall, these theses are:

- A suitably modified version of *holism* can have an important role to play in our best theory of concepts.
- The most plausible version of the publicity principle is the *contextualist* one developed in chapter 3.

As we will see, once we examine the range of possible ways to solve the (PUB)-(FC) tension these theses become hard to resist, since:

- On *both* of our solutions, a modified holistic condition will look like the best way to satisfy (FC).
- The least problematic of our solutions seems to *require* a contextualist version of publicity.

I will then conclude the chapter by sketching my own picture of concept individuation and possession; this, I think, is the picture of concepts that emerges as a clear winner in light of our arguments in chapters 1-5.

2. The Implications Of (FC)

2.1. What (FC) Demands

In this section, I will analyze the Fregean constraint and its implications in more detail. In particular, I will argue that (FC) entails a certain *necessary condition* for concept sameness and that this condition is a *very fine-grained one*.

Here is how I propose to proceed. (FC) holds that, whenever two concepts are involved in the contradictory beliefs of a rational subject, those concepts must be distinct; this, recall, is a necessary assumption if we want to explain the subject's rationality. But now suppose we can show that, whenever two concepts C_1-C_2 have certain properties (e.g. whenever they stand in a certain relation r), C_1-C_2 will be involved in contradictory beliefs. It will then follow immediately from (FC) that, if two concepts C_1-C_2 stand in r , then they are distinct concepts: a sufficient condition for concept difference, which is of course equivalent to a necessary condition for concept sameness²²². By finding out the conditions under which two concepts will be involved in contradictory beliefs, then, we will also find out what conditions for concept sameness are entailed by (FC); we will find out, roughly speaking, "how finely" (FC) requires us to "cut" concepts.

In this section I will argue that, by standard RTM principles, whenever two concepts stand in certain *inferential connections* they will also be involved in contradictory beliefs. I will then go on to examine (sect. 2.2) the condition on concept sameness that results from this; as we will see, that condition is an extremely fine-grained one, something we will have to keep in mind when developing our individuation criterion in the last section of this chapter.

As we have seen, an important tradition within RTM holds that, in order to explain how someone can rationally have contradictory *de re* beliefs about some object, we must assume that the concepts involved in those beliefs are distinct. This commits the RTM theorist to the constraint we examined in chapter 4:

The Fregean Constraint (FC): If two concepts C_1-C_2 are involved in the contradictory beliefs of a rational subject S at time t about some object x , then C_1-C_2 are not the same concept.

Now, what we are looking for are the conditions under which (FC) will entail that two concepts C_1-C_2 are distinct. To find out what these conditions are, I will analyze a certain hypothetical case. In fact, the case is already familiar to the reader, since it is simply a

²²² (It follows from our sufficient condition that *not standing in r* is *necessary* for two concepts C_1-C_2 to be the *same* concept).

general “template” for specific Frege cases such as Lois’, Edmund’s, Peter’s and so on: each of these cases satisfies the assumptions holding in our case and can therefore be considered as one of its instances.

Suppose there are two concepts C_1 and C_2 such that:

- 1) C_1 and C_2 are had by a rational subject S at time t .
- 2) C_1 and C_2 refer to x .
- 3) C_1 has [IS F], but not [IS NOT F], among its inferential connections.
- 4) C_2 has [IS NOT F], but not [IS F], among its inferential connections.

Assumptions (1)-(2) do not require much clarification. What we are imagining is a subject like Lois or Edmund, who is normal under all relevant respects, appears rational in his doxastic behavior, and has two coreferential concepts for some object x at some time t . Assumptions (3)-(4) concern the specific inferential connections in which S ’s two concepts stand. What we are imagining is a subject who:

- Is disposed to infer from X IS IDENTICAL TO C_1 to X IS F; that is, he accepts the thought C_1 IS F²²³.

(On the other hand, he is *not* disposed to infer from X IS IDENTICAL TO C_1 to X IS NOT F; that is, he does not *also* accept the thought C_1 IS NOT F. If he did, he would arguably be *irrational*).

- Is disposed to infer from X IS IDENTICAL TO C_2 to X IS NOT F; that is, he accepts the thought C_2 IS NOT F.

(On the other hand, he is *not* disposed to infer from X IS IDENTICAL TO C_2 to X IS F; that is, he does not *also* accept the thought C_2 IS F. Again, he would arguably be irrational if he did).

S ’s situation is, of course, perfectly familiar. S is just like Lois, who has:

- A concept SUPERMAN having [CAN FLY] (but not [CANNOT FLY]) among its inferential connections;

²²³ By standard RTM principles, whenever we say that a concept “stands in a certain inferential connection” we can also say that the owner of the concept accepts the corresponding thought: to say that S is disposed to infer from X IS IDENTICAL TO C_1 to X IS F is just to say that S accepts the thought C_1 IS F.

- A concept CLARK having [CANNOT FLY] (but not [CAN FLY]) among its inferential connections.

All other subjects in Frege cases also fall under the general template constituted by S's case: cf. for instance Edmund, whose SUPERMAN₁ is connected to [CAPTURED THE ROBBER] while his SUPERMAN₂ is connected to [DID NOT CAPTURE THE ROBBER]. If you want, you can also think of the relevant concepts as “mental files” storing information about their referents²²⁴. Lois' SUPERMAN-file contains [CAN FLY] but not [CANNOT FLY]; vice versa, her CLARK-file contains [CANNOT FLY] but not [CAN FLY]. Similarly, S's C₁-file contains [IS F] but not [IS NOT F], and his C₂-file contains [IS NOT F] but not [IS F]. Again, if one the two files contained both [IS F] and [IS NOT F], S would arguably be irrational; that would amount to accepting both C IS F and C IS NOT F at the very same time.

Assuming that (1)-(4) are true of two concepts C₁-C₂, our driving question will be very simple: *given (FC), are C₁ and C₂ different concepts?* If they are, we have found out a sufficient condition for concept difference: (FC) entails that, if (1)-(4) are true of two concepts, then those concepts are different.

Now, if we took “different concepts” to mean *numerically distinct* concepts, the answer to our question would be trivial. By the indiscernibility of identicals, if two objects have different properties then they are numerically distinct. But, by hypothesis, our two concepts have different inferential connections: for instance, C₁ is inferentially connected to [IS F] while C₂ is not. Therefore, our two concepts have different properties and must be numerically distinct. So, if by “different concepts” we mean *numerically different* concepts, C₁-C₂ will clearly count as different; indeed, this follows from (1)-(4) even if we don't assume (FC) at all!

Suppose on the other hand that we take “different concepts” to mean not *numerically distinct* but, for instance, *belonging to distinct concept types*. Under this assumption, it's clearly not trivial to show that (FC) requires C₁-C₂ to be “different concepts”: C₁-C₂ might well belong to the *same* concept type even if they are *numerically* different. Here is a different way to put the same point. A non-Fregean ontologist who was looking for a set of typing conditions²²⁵ would accept as trivial the claim that two token concepts satisfying (1)-(4) must be numerically distinct. At the same time, however, he certainly won't find it trivial that, if (FC) is true, then two token concepts satisfying (1)-(4) must also belong to different concept types. After all, it's at least not obvious that a set of typing conditions which counted those two tokens as falling under the same type would be inconsistent with (FC)²²⁶.

²²⁴ V. Recanati (1993, 2009), Perry (2000).

²²⁵ V. e.g. the typing criteria by Fodor and Schneider discussed in ch. 4.

²²⁶ Objection: “By hypothesis, C₁-C₂ differ in some of their properties. Therefore, we can trivially show that they belong to distinct types once we identify type concepts with *classes* of concepts, as you did in ch. 3 (sect. 4.1)”. Reply: non-Fregeans like Fodor and Schneider think a concept can only belong to *one*

In what follows, then, my aim is not just to prove the (trivial) claim that two concepts satisfying (1)-(4) will be “different concepts” in the numerical-difference sense. What I want to defend is the (non-trivial) claim that, if (FC) is true, then C_1 - C_2 are “different concepts” *even if we take the phrase to express difference in type*; a consequence that, as we will see later, has very important implications for our general goal of finding a satisfactory individuation criterion for concepts.

My argument is very simple and rests on a single premise:

- Since S’s concept C_1 has [IS F] among its inferential connections, C_1 is *involved* in S’s belief, of x , that x is F.
- And since S’s concept C_2 has [IS NOT F] among its inferential connections, C_2 is *involved* in S’s belief, of x , that x is not-F.

I take this premise to be entirely uncontroversial for anyone who endorses the standard RTM framework. When introducing (FC)²²⁷, I defined involvement” as follows: if S has a *de re* belief about some x in virtue of accepting a thought that is partially constituted by some concept C, then C is involved in S’s belief. (So SUPERMAN-CLARK are involved in Lois’ contradictory beliefs about Superman/Clark). Given this definition, it’s easy to show that the above premise follows from RTM *plus* our description of the case:

- Since S’s C_1 has [IS F] among its inferential connections, S accepts the thought C_1 IS F; and since S’s C_2 has [IS NOT F] among its inferential connections, S accepts the thought C_2 IS NOT F.

(On RTM, to accept the thought C_1 IS F *just is* to have a concept C_1 which is inferentially connected to [IS F]; *mutatis mutandis* for C_2).

- But to accept C_1 IS F is to stand in the belief-relation to C_1 IS F, and to accept C_2 IS NOT F is to stand in the belief-relation to C_2 IS NOT F²²⁸.
- Finally, according to RTM²²⁹, to stand in the belief-relation to C_1 IS F is to believe, of the referent of C_1 , that it is F. And to stand in the belief-relation to C_2 IS NOT F is to believe, of the referent of C_2 , that it is not-F.

type at a time. This becomes apparent once we note that, according to these theorists, there is only *one* correct typing criterion for concepts (cf. ch. 4, sect. 3.1). Obviously, then, these theorists do *not* identify type concepts with classes (a token concept can belong to more than one class at the same time); therefore, my argument will not be trivial for them. I will criticize their view of type concepts in sect. 4.3 *infra*, but I’m not presupposing my theory of types at this stage.

²²⁷ Ch. 4, sect. 2.2.

²²⁸ “Accepting a thought” was introduced as an abbreviation for “standing in the belief-relation to a thought” in sect. 2.1, ch. 2.

- So C_1 is involved in S's belief, of x , that x is F, and C_2 is involved in S's belief, of x , that x is not-F.

Clearly, no one who accepts RTM, and particularly a proponent of (FC), could reject this conclusion. Given RTM, it's almost true by definition that, if [IS F] is among the inferential connections of some concept C, then the owner of C believes of C's referent that it is F; *mutatis mutandis* for [IS NOT F]. Our argument will then go as follows:

- We have just shown that (by assumptions (2) to (4)) C_1 is involved in S's belief, of x , that x is F, and C_2 is involved in S's belief, of x , that x is not-F.
- So C_1-C_2 are involved in the contradictory beliefs of a subject S at time t about some object x .
- By assumption (1), S is rational.
- So C_1-C_2 are involved in the contradictory beliefs of a rational subject S at time t about some object x .
- Therefore, it follows from (FC) that C_1-C_2 are not the same concept.
- Conclusion: *(FC) entails that, if assumptions (1)-(4) are true of two concepts C_1-C_2 , then C_1-C_2 are not the same concept.*

We have found a sufficient condition for concept difference entailed by (FC). In less formal terms, how does the argument go? Let's say that when two concepts satisfy assumptions (3)-(4) they have “*contradictory inferential connections*” (C_1 has [IS F] but not [IS NOT F], C_2 has [IS NOT F] but not [IS F], and nothing can be both F and not-F, so their inferential connections are “*contradictory*”). As we have seen, it's almost true by definition that, when two coreferential concepts have contradictory connections, they will be involved in contradictory *de re* beliefs. This will raise our familiar Rationality Problem: how can someone rationally have contradictory beliefs? According to (FC), answering this question requires assuming that the concepts involved in those beliefs are *distinct*: if we didn't make this assumption, we would have no explanation of the subject's rationality. So (FC) entails that two coreferential concepts with contradictory connections must be *different concepts*. If they weren't, there would be no way to explain the rationality of their owner.

Crucially, notice that this is true *whatever kind* of difference relation is expressed by “*different concepts*” here. If the relation expressed is one of numerical difference, then it will be trivial to show that concepts standing in contradictory connections are

²²⁹ V. RTM's principle (I) in sect. 2.1, ch. 2.

“different”. Even if the relation expressed is one of type-difference, however, (FC) will still entail that concepts with contradictory connections are “different”: that is, it will entail that two such concepts must belong to distinct types. And this is not trivial.

2.2. A Fine-Grained Condition

Let’s summarize our argument so far. I have argued that (FC) entails the following conditional:

(C) If assumptions (1)-(4) are true of two concepts C_1-C_2 , then C_1-C_2 are different concepts.

Assumptions (1)-(4) were:

- 1) C_1-C_2 are had by a rational subject S at time t .
- 2) C_1-C_2 refer to x .
- 3) C_1 has [IS F], but not [IS NOT F], among its inferential connections.
- 4) C_2 has [IS NOT F], but not [IS F], among its inferential connections.

We can therefore restate (C) as follows:

(C) If two concepts C_1-C_2 are had by the same rational subject S at the same time t (= (1)) and they are coreferential (= (2)), then if they have contradictory inferential connections (= (3)-(4)) they are different concepts²³⁰.

Now, (C) states a sufficient condition for concept difference. But, of course, a sufficient condition for concept difference is a necessary condition for concept sameness. So let’s restate (C) as a necessary condition:

(C) If two concepts C_1-C_2 are had by the same rational subject S at the same time t and they are coreferential, then **if they are the same concept they do not have** contradictory inferential connections.

In slightly more formal terms:

(C) (two concepts C_1-C_2 are had by the same rational subject S at the same time t and they are coreferential) \rightarrow (they are the same concept \rightarrow they do not have contradictory inferential connections)

²³⁰ Recall that “having contradictory inferential connections” has been *defined* as “satisfying assumptions (3)-(4)” (v. end of sect. 2.1).

Which is of course equivalent to:

(C) (two concepts C_1-C_2 are had by the same rational subject S at the same time t and they are coreferential **and they are the same concept**) → (they **do not have** contradictory inferential connections)

In sum: our argument in the last section shows that (FC) entails a certain *necessary condition on concept sameness*, which is stated by our conditional (C). Before we analyze the implications this has for our problem, notice that (C) is a “*restricted*” condition on concept sameness. (C) *only* applies to pairs of concepts had by the *same* subject at the *same* time, and *only* if the subject is rational and the concepts coreferential. Suppose two concepts C_1-C_2 fail to satisfy one of these assumptions; for instance, suppose they don’t belong to the same subject. Clearly, it won’t follow from (C) that C_1-C_2 are different, even if they happen to stand in contradictory connections; for all (C) says, C_1-C_2 might still be the same concept. This point can also be seen by noting that (C) is equivalent to:

(C) For any two coreferential concepts C_1-C_2 had by the same rational subject S at the same time t : (C_1-C_2 are the same concept → they do not have contradictory inferential connections)

Where the quantifier is *restricted* to a specific set of concepts, instead of ranging over all concepts as in unrestricted conditions. In conclusion, (C) does impose a necessary condition on concept sameness (not having contradictory connections), but only on coreferential concepts had by the same subject at the same time: using our terminology from the last chapter, (C) is a “*restricted*” condition.

An important *caveat*: to simplify our discussion, from now on I will *ignore* two of the restrictions built into (C). These are the restrictions which require C_1-C_2 to be *coreferential* and belong to a *rational* subject. I will therefore treat (C) as equivalent to:

(C) (two concepts C_1-C_2 are had by the **same subject S** at the **same time t** and they are the same concept) → (they do not have contradictory inferential connections)

The simplified version is still restricted, as it only applies to pairs of concepts had by the same subject (it only applies *intrapersonally*) at the same time (it only applies *synchronously*). While the simplified version does not in fact follow from the original one, the restriction on subjects/times is the only one that really matters for our purposes, and all of my arguments in what follows would go through *a fortiori* on the original version. Since the new version simplifies things significantly, I will stick to it in what follows. (For now, I must simply ask the reader to trust me on this score; once my arguments are on the table, he will be able to check for himself whether this simplification is legitimate or not).

Having clarified (C)'s status, we can finally see its relevance for our problem. I have shown that there is a condition on concept individuation which:

- Follows from an *important constraint* on a theory of concepts, namely (FC);
- Only applies to *certain concepts*, i.e. those had by a single subject at the same time;
- “Slices” concepts *very finely*.

Each of these points should be easy to see at this stage. Regarding the first point, my argument has shown that (FC) entails condition (C) on concept sameness. Regarding the second, we have seen that this condition only applies to concepts had by the same subject at the same time, and holds that two of *these* concepts are identical *only if* they don't have any contradictory connections.

The third thing to note is that our condition is an *extremely fine-grained* one. If (C) is true, then the presence of *any* contradictory connections is enough for certain concepts to be different. (FC) thus requires individuating concepts had by a single subject at the same time in a very fine way. For any two such concepts, it *doesn't matter* whether they have same reference, same origin, how many inferential connections they have in common, and so on; as long as there is *even only one* contradictory connection in their inferential roles, that's enough for them to be distinct. In a nutshell: given (FC), our *intrapersonal* and *synchronic* individuation criteria for concepts must be extremely fine-grained.

To see just how fine-grained our necessary condition is, it will be useful to see what happens once we eliminate all restrictions to subjects/times in (C). What would happen if we took (C) to apply to *all concepts*, not just those had by a single subject at the same time? The following is the unrestricted version of (C):

(C*) Two concepts C_1 - C_2 are the same concept \rightarrow they do not have contradictory inferential connections

(C*) simply holds that, *for any two concepts*, if some of their inferential connections are contradictory then the two concepts are distinct. But this cuts concepts too thin: the unrestricted version of (C) is straightforwardly incompatible with (PUB). The following is a consequence of (C*): whenever I ascribe F to x and you ascribe not-F to x , our two concepts of x are not the same concept. So suppose I believe that tomatoes were present in Europe before year 1492, while you believe they were not. Our two concepts for tomatoes have contradictory connections: my concept $TOMATO_1$ is connected to [PRESENT IN EUROPE BEFORE 1492], while your concept $TOMATO_2$ is connected to [NOT PRESENT IN EUROPE BEFORE 1492]. According to (C*), this suffices for our two concepts to be distinct. Even worse, the same will be true of two time-slices of the same person: when I learn that tomatoes were not present in Europe before 1492, my future time-slice will have a concept $TOMATO_2$ which is not the same concept as my current concept

TOMATO₁. But, of course, publicity entails that you and I have a concept for tomatoes in common; similarly for me and my future, more knowledgeable time-slice. Since (C*) entails that our concepts for tomatoes are different, the unrestricted version of (C) is straightforwardly incompatible with (PUB). (Of course, the same problem will arise for many other shared concepts like WATER, DOG, SUPERMAN etc....).

Of course, we must keep in mind that (FC) does *not* entail the unrestricted (C*), but only the restricted (C); and (C) is *consistent* with publicity. ((C) only applies to concepts had by a single subject, while (PUB) only applies to concepts had by different subjects²³¹. So (C) is guaranteed to be compatible with (PUB). We will go back to this point at length in sect. 4.1 *infra*). Therefore, the moral of our discussion is not that (FC) is inconsistent with (PUB). It is, rather, that (FC) entails a very fine-grained sameness condition for a certain class of concepts, i.e. those had by a single subject, at a specific time. That condition is so fine-grained that, if applied at the interpersonal/diachronic level, it would clearly make it too hard for different thinkers to have the same concepts.

Let's take stock. I have argued that (FC) entails a certain necessary condition (C). Once we accept (FC), this means that (C) must be included in our individuation criterion for concepts (together with other necessary/sufficient conditions for concept sameness, of course). More specifically, (FC) requires that we individuate a specific class of concepts (those had by the same subject at the same time) in a very fine-grained way: we must treat such concepts as different whenever some of their inferential connections are contradictory. As we will see in section 4, this has crucial consequences: in particular, it helps make other fine-grained sameness conditions such as holism a lot more plausible, given that our individuation criterion will have to include a fine-grained condition *anyway*. Our results in this section will thus have an important role in establishing one of my main theses in this chapter.

Before we discuss this point in more detail, however, I want to go back to the problem raised in the last chapter: I believe we are now in a better position to see why unrestricted individuation criteria tend to run into trouble with publicity. This is what the next section is about.

²³¹ As we have already noted on several occasions, different time-slices of the same person count as different subjects for our purposes (cf. e.g. ch. 2, sect. 2.2). For this reason, I will often use the expression “different subjects” to cover different time-slices of the same person, as well as different people; and I will use the expression “same subject” to mean not only “same person”, but also “same time-slice of the same person”.

3. The Source Of The Problem

In chapter 4, we discussed various (unrestricted) individuation criteria that have been put forward by some of the major theories of concepts on the market. We noted that these criteria seem to face a dilemma: some of them simply fail to satisfy (FC) in Frege cases such as Edmund's, while others do satisfy (FC) but face the following problematic consequences:

- a) They are *straightforwardly inconsistent* with (PUB) (cf. condition (b) (originalism); condition (c) in its holistic version; condition (d) in both its physicalist and its holistic versions).
- b) *Or*, at any rate, they have the potentially problematic consequence that many concepts involved in Frege cases (e.g. Edmund's SUPERMAN₁-SUPERMAN₂) are *not* widely shared across ordinary intentional subjects (cf. the “modified” criteria discussed in ch. 4, sect. 4.3).

Now, one might reasonably complain that I have only examined some possible unrestricted conditions: no doubt, other candidates are available, so why think that every unrestricted condition will have one of the above consequences? This section will try to answer this question: more specifically, my aim is to:

- Explain *why* the problem arises, i.e. why unrestricted conditions that satisfy (FC) have the consequences described above.
- Provide principled reasons to think that *any* such condition will have the same consequences: we can't solve the problem by simply looking for a more sophisticated account.

To see why our problem arises, it will be useful to reflect on what an individuation criterion that satisfies (FC) will look like. Consider the set of all cases in which two concepts are had by the same (rational²³²) subject and are involved in contradictory beliefs: for any such case, an individuation criterion that satisfies (FC) will have to include a suitable necessary condition. In particular, take an arbitrary case *c* in which two concepts C_X-C_Y are involved in contradictory beliefs. For every such case, our individuation criterion will include a specific necessary condition of the following form:

(K1_c) A concept C is the same concept as C_X → concept C has property/properties [...]

²³² As usual, I will omit this qualification in what follows.

If our criterion does satisfy (FC), then ($K1_c$) will include on its right hand side some property F which C_Y does *not* have: it will then follow from ($K1_c$) that C_Y is *different* from C_X , which is just what's needed to satisfy (FC) in that particular case. (As usual, ($K1_c$) might simply follow from a more general sameness condition, as with all the major candidates examined in ch. 4. Alternatively, one could provide a specific condition for every possible Frege case, where the conditions provided might vary from case to case²³³). Going back to Edmund's example: if a criterion satisfies (FC), then it must include a necessary condition like:

($K1_{Edmund}$) A concept C is the same concept as SUPERMAN₁ → concept C has property/properties [...]

We will then have to fill in the brackets with some property F which SUPERMAN₁ has and SUPERMAN₂ lacks²³⁴.

In sum: for *every* Frege case c in which two concepts C_X-C_Y are involved in contradictory beliefs, there will have to be some condition ($K1_c$) specifying, on its right hand side, some property F which C_X has, and C_Y lacks. How does this help explain the problems that unrestricted criteria seem to face? Suppose that, in order to satisfy (FC) in a case c , we include in our criterion a condition ($K1_c$) which is *not* restricted to concepts had by the same subject. It will then follow from ($K1_c$) that *any* concept which lacks F is different from C_X , no matter whether it's had by the same subject as C_X (like C_Y) or by a different subject. Consider for instance the modified localist account from chapter 4 (sect. 4.3), which puts forward the unrestricted condition:

($K1_{Edmund}$) A concept C is the same concept as SUPERMAN₁ → concept C has same inferential role as SUPERMAN₁

Given how the localist construes SUPERMAN₁'s role, this condition is equivalent to:

($K1_{Edmund}$) A concept C is the same concept as SUPERMAN₁ → concept C is inferentially connected to [CALLED "SUPERMAN", WEARS A RED CAPE, ... DID NOT SAVE THE KITTEN, CAPTURED THE ROBBER]

²³³ V. ch. 4, sect. 3.2 for more details.

²³⁴ As usual, the condition provided will be interpreted differently depending on our background ontology. On a Fregean ontology, ($K1_{Edmund}$) will amount to:

($K1_{Edmund}$) A concept C is numerically identical to SUPERMAN₁ → concept C has property/properties [...]

On a non-Fregean ontology:

($K1_{Edmund}$) A (token) concept C belongs to the same type as (token) concept SUPERMAN₁ → concept C has property/properties [...]

It thus follows from (K1_{Edmund}) that *any* concept which is *not* connected to [CAPTURED THE ROBBER] is different from SUPERMAN₁. This, of course, results in SUPERMAN₁ *not* being shared by ordinary subjects like Verna²³⁵, who just don't have a concept meeting that condition. Similar consequences ensue with the other unrestricted criteria we examined, resulting *either* in a violation of publicity or, at any rate, in the non-shareability of those concepts which are involved in a subject's contradictory beliefs.

We can now see *why* unrestricted criteria have this consequence. Consider the general features of the cases we are considering. In each of them, two concepts C_X-C_Y are involved in contradictory beliefs, and we must find a property F which only C_X has, so that we can include it in (K1_c). This property could be: being inferentially connected to [CAPTURED THE ROBBER]; having a certain origin o; having a certain set s of physical-neurological properties; and so on. Be that as it may, it is very likely that C_Y will *not* be the only concept lacking F. Presumably, many concepts had by other subjects will also fail to instantiate that property; indeed, there will be several subjects who don't have *any* concepts with F (cf. Verna again). *Because our condition is unrestricted*, then, it will follow that the subjects in question don't have C_X. (If our condition was restricted, i.e. if it only applied to concepts belonging to the same subject as C_X, the problem would not arise: concepts belonging to other subjects could still be the same as C_X even if they lacked F, since (K1_c) would not apply to them).

Here, then, is my explanation of the problem from chapter 4 in brief:

- a) For every two concepts C_X-C_Y involved in a Frege case, an *unrestricted* condition (K1_c) that satisfies (FC) will specify a property F which C_X has, and C_Y lacks.
- b) For many Frege cases, concepts had by *other* intentional subjects are *also* likely to lack F.
- c) *Because* (K1_c) is unrestricted, it will follow that these subjects do not have a concept that is the same as C_X.
- d) Conclusion: for many Frege cases, an unrestricted condition that satisfies (FC) will also entail that the concepts involved in those cases are not shared.

Which, of course, is precisely what happens with the all the conditions examined in chapter 4.

While steps (a) and (c) of the explanation are entirely uncontroversial, one might wonder what kind of argument can be offered for the crucial premise (b). Why think that F (= the property which differentiates the concepts involved in a Frege case) will not be widely shared? Answer: it's hard to imagine what kind of property F might be, such that it could play these two roles:

²³⁵ Cf. ch. 4, sect. 4.2.

- Being instantiated by C_X , but not by C_Y ;
- Being widely instantiated across concepts had by other intentional subjects.

If C_Y lacks F, why think that this will be a unique feature of that concept? What is special about C_Y , so that it fails to instantiate a property which concepts had by other subjects are guaranteed to have? To see the point more clearly, imagine what would be required for F to be widely shared in *all* Frege cases. The following would then have to obtain: for every Frege case, there is a property F which is *not* instantiated by one of the concepts involved, but which is *guaranteed* to be shared by concepts had by other subjects. Clearly, a principled argument would be needed to show that, for every Frege case, there is always going to be such a property. But, first, it's hard to imagine what that argument might look like and, second, our previous discussion gives us good reasons to think that such an argument won't be available. In a paradigmatic Frege case such as Edmund's, all the candidate properties we examined (*having such-and-such inferential role*; *having such-and-such origin*; etc...) fail to satisfy one of our two *desiderata*: to the extent that $SUPERMAN_2$ lacks one of them, concepts had by other intentional subjects also do. Since it seems possible to devise countless other cases that are structurally analogous to Edmund's, we have all reasons to believe that, in many Frege cases, F is not going to be widely shared.

Now, there clearly are some properties which could play the two roles above. For instance, suppose that:

$F = \text{being connected to } [\text{CAPTURED THE ROBBER}] \text{ or belonging to a different subject from } SUPERMAN_1$

This property is indeed special: Edmund's $SUPERMAN_1$ has it, his $SUPERMAN_2$ lacks it, and it is guaranteed to be shared by concepts had by other subjects (e.g. Verna). This is because, of course, these concepts don't belong to the same subject as $SUPERMAN_1$, so they all instantiate our disjunctive property F in virtue of satisfying its second disjunct. Problem is, F would make our necessary condition ($K1_{\text{Edmund}}$) *restricted*. Once we fill in the brackets with F, the condition becomes:

($K1_{\text{Edmund}}$) A concept C is the same concept as $SUPERMAN_1 \rightarrow$ concept C is inferentially connected to [CAPTURED THE ROBBER] **or** it belongs to a **different subject** from $SUPERMAN_1$

Which is equivalent to:

($K1_{\text{Edmund}}$) A concept C belongs to the **same subject** as $SUPERMAN_1$ **and** it is the same concept as $SUPERMAN_1 \rightarrow$ concept C is inferentially connected to [CAPTURED THE ROBBER]

Which is a condition that only applies to concepts had by the same subject as SUPERMAN₁, i.e. a restricted sameness condition²³⁶. I am of course happy to grant that such a condition would satisfy (FC) without entailing that SUPERMAN₁ is not shared. My claim, however, is that *unrestricted* conditions will cause SUPERMAN₁ to be non-shared, and such conditions cannot help themselves to disjunctive properties like the above.

I have tried to explain why unrestricted criteria that satisfy (FC) have the consequence that concepts involved in Frege cases are not shared, thus running into troubles with publicity. This shows that the problem examined in chapter 4 was not an accident: there are principled reasons to think that more sophisticated criteria will have exactly the same consequence. Bottom line: *if* we try to satisfy (FC) through an unrestricted condition, we will have to accept that many concepts are not shared by ordinary intentional subjects. With this conclusion in mind, we can finally start looking for a solution to our problem.

²³⁶ To see this more clearly, notice that the condition is equivalent to:

(K1_{Edmund}) For every concept C that belongs to the **same subject** as SUPERMAN₁: (C is the same concept as SUPERMAN₁ → C is inferentially connected to [CAPTURED THE ROBBER])

Where the quantifier is *restricted* to concepts had by the same subject as SUPERMAN₁. (V. sect. 2.2 *supra* for more details about restricted sameness conditions).

4. Concepts: A Pluralist-Contextualist Picture

4.1. Two Solutions

Our discussion in chapter 4 began with two constraints and a problem: every unrestricted individuation criterion that could satisfy (FC) also seemed to violate (PUB). Several major theories of concepts were discussed, and none of them appeared to escape the dilemma (ch. 4, sect. 4.2). Then, we considered a possible way to “patch” those theories (ch. 4, sect. 4.3): one could offer two different sets of unrestricted conditions, a more fine-grained one for the non-shared concepts involved in Frege cases, and a more coarse-grained one for the shared concepts covered by (PUB). While this still entails that many concepts are not shared, we left open whether this consequence could after all be accepted as compatible with publicity. It’s now time to go back to our “patch”, reassess it and compare it with some alternative strategies. Our first goal (sect. 4.1) will then be:

Goal n. 1: Examine various possible solutions to our problem, that of finding an individuation criterion for concepts that satisfies (FC) without violating (PUB).

I will first develop in more detail the solution sketched in chapter 4; then, I will compare it with one main alternative, which consists in offering a *restricted* individuation criterion that can satisfy (FC) while avoiding the commitment to unshared concepts²³⁷.

Having briefly compared our two solutions, I will proceed to my second goal (sects. 4.2-4.3):

Goal n. 2: Use my solution to the original problem to defend two main theses: that a (suitably modified) form of holism should “have a place” in a theory of concepts, and that the best version of the publicity principle is the “contextualist” one developed in chapter 3.

More specifically, I will argue that on *both* of our solutions we will have excellent reasons to include a modified holistic condition in our individuation criterion, and that on *the most plausible* of those solutions we will also have excellent reasons to adopt a contextualist version of publicity.

Having defended my main theses, I will conclude on my final goal (sect. 4.3):

²³⁷ While other solutions might be available, these appear to be the most promising ones in light of our previous discussion.

Goal n. 3: Sketch the general picture of concept individuation and possession that best fits with the general results of our discussion so far.

I believe our arguments in chapters 1-5 support a specific picture of concept individuation/possession, one which emerges as overwhelmingly plausible in light of the various conclusions we have reached. That picture integrates all the main theses I defended along the way and, for reasons that will be obvious later, it can aptly be called a “pluralist-contextualist” theory of concepts.

Let us now turn to our first goal and examine two possible ways to develop an individuation criterion for concepts that satisfies (FC) without violating (PUB).

First Solution: Living With Unshared Concepts

Our first solution was sketched in chapter 4 (sect. 4.3), where I noted how some of the conditions we examined could be modified so as to satisfy (FC) without immediately violating (PUB). There, I focused on a modified version of *localism* to give the reader an idea of how the solution might go. We can now describe the solution in more general terms and then proceed to assess its merits.

The solution goes as follows. First, we provide an *unrestricted* necessary condition in order to satisfy (FC): for every case in which two concepts C_X-C_Y are involved in contradictory beliefs, our individuation criterion will then include a necessary condition of the form:

(K1_c) A concept C is the same concept as $C_X \rightarrow$ concept C has property/properties [...]

Where, of course, C_Y will *lack* whatever property F appears in brackets (analogous unrestricted conditions will then be provided for every other Frege case). Here is a familiar example of the kind of condition I have in mind: a “modified localist” view (ch. 4, sect. 4.3) will satisfy (FC) in Edmund’s case through the unrestricted:

(K1_{Edmund}) A concept C is the same concept as SUPERMAN₁ \rightarrow concept C has same inferential role as SUPERMAN₁

Where SUPERMAN₁’s inferential role includes the inference to [CAPTURED THE ROBBER], so that the account can yield the right predictions in Edmund’s case.

Now, we have seen that there are both “inductive” (ch. 4) and “principled” (sect. 3 *supra*) reasons to think that, for several Frege cases, our unrestricted condition will entail that C_X-C_Y are *not* shared. For instance, all the individuation criteria examined in chapter 4 (including modified localism) had precisely this consequence in Edmund and Verna’s case: once we decide to satisfy (FC) through an unrestricted criterion, we must

grant that the concepts involved in Edmund's beliefs ($SUPERMAN_1$ - $SUPERMAN_2$) are *not* shared by subjects like Verna, even though these subjects can engage with Edmund in Superman-related communication, fall under the same Superman-related generalizations etc... *Mutatis mutandis* for other Frege cases that have the same features.

In sum: if the criterion we use to satisfy (FC) is unrestricted, we must concede the following: *those concepts which are involved in contradictory beliefs* (e.g. $SUPERMAN_1$ - $SUPERMAN_2$) *are often not shared by other intentional subjects*. We will then have to show that this consequence of our unrestricted criterion is in fact compatible with publicity; for instance, we must show that publicity *doesn't* require Edmund's $SUPERMAN_1$ to be shared by Verna. Fortunately, there is a strategy we can use for this purpose²³⁸. As usual, this strategy will have to be declined differently, depending on whether we adopt a Fregean or a non-Fregean ontology; I will consider each of them in turn²³⁹.

A Fregean might hold that publicity does *not* require Verna to have a concept which is numerically identical to Edmund's $SUPERMAN_1$: Edmund and Verna can successfully communicate, agree/disagree about Superman etc... as long as they have *some* concept of Superman in common. For instance, there might be a concept $SUPERMAN$ which is less finely individuated than Edmund's $SUPERMAN_1$, and which both subjects have. (This concept might e.g. be individuated by the standard Superman-ish role, in which case Verna and Edmund would both have it). Bottom line: publicity is not violated, even though some of Edmund's Superman-concepts are not shared by Verna.

A non-Fregean ontologist has two strategies available (both are structurally similar to the Fregean one). First, he could interpret (K1c) as a condition for *numerical identity* of token concepts, rather than a condition for type identity. For instance, our modified localist view could hold that a token concept c is numerically identical to Edmund's token concept $SUPERMAN_1$ only if they have same inferential role. This would satisfy (FC), since $SUPERMAN_2$ would then be numerically distinct from $SUPERMAN_1$. Of course, it would also follow that none of Verna's token concepts is numerically identical with Edmund's $SUPERMAN_1$. This is unproblematic, however, since Verna's concepts can still belong to the same *type* as $SUPERMAN_1$. For instance, we could adopt a coarse-grained typing condition like: two token concepts belong to the type concept $SUPERMAN$ iff they have the standard Superman-ish role. Verna and Edmund would then have numerically distinct Superman-concepts belonging to the same type, and publicity would be respected.

²³⁸ The reader will note that this is the strategy I employed in ch. 3 (sect. 4.1) while developing my “modified holistic” view; more on this soon.

²³⁹ I will ignore those solutions that couldn't be accepted within a certain ontology or that simply appear downright implausible.

Alternatively, the non-Fregean might decide to satisfy (FC) through a condition for *type identity*, while holding that a concept can belong to *multiple types*²⁴⁰. For instance, he might put forward the following condition:

(K1_{Edmund}) A token concept C **belongs to the same type concept T₁** as SUPERMAN₁
→ concept C has same inferential role as SUPERMAN₁

The idea behind this condition is that there is a type-concept T₁ whose membership conditions are extremely fine-grained: a token concept must have SUPERMAN₁'s inferential role in order to belong to T₁. (As usual, we will assume [CAPTURED THE ROBBER] is part of SUPERMAN₁'s role). Again, our condition makes the right predictions about Edmund's concepts, since only one of them belongs to T₁: the two concepts are type-different. Of course, the condition also entails that none of Verna's concepts belongs to T₁. If a concept can belong to multiple types, however, this need not violate publicity, since there might be a further type T to which *both* Edmund and Verna's concepts belong. As usual, T's membership conditions will be coarse-grained: for instance, T might be simply individuated by the standard Superman-ish role, in which case our two subjects will both have token concepts belonging to it. Even though T₁ is not shared, then, publicity is respected because T is.

Clearly, these three versions of our solution closely parallel each other. Setting aside ontological subtleties, they all make the same claim: each of us has a stock of shared concepts, and a stock of unshared ones. The unshared ones are those involved in the Frege cases that arise all the time in our cognitive lives; the shared ones are those that account for our ability to communicate, agree/disagree, fall under the same generalizations and so on. All publicity requires is that we share *some* of our concepts with other intentional subjects; we are therefore free to provide unrestricted criteria for concept sameness, even though this will have the consequence that many concepts (such as Edmund's SUPERMAN₁) aren't public.

Second Solution: Restricted Individuation Conditions

Let us now consider a radically different solution to our problem. Since we started looking for an individuation criterion that could satisfy (FC), we only considered criteria of a certain type, which I have called “unrestricted”; as we have seen in chapter 4, all the leading theories of concepts currently on the market have proposed individuation criteria of this kind. Given a Frege case involving two concepts C_X-C_Y, these criteria hold that, in order for *any* concept to be the same as C_X, that concept must have a certain property (e.g. being inferentially connected to [CAPTURED THE ROBBER]);

²⁴⁰ V. ch. 3 (sect. 4.1) for discussion of some possible objections against the appeal to multiple types.

having C_x 's origin; having C_x 's physical-neurological properties; and so on). The sameness condition provided will therefore have the following form:

(K1_c) A concept C is the same concept as $C_x \rightarrow$ concept C has property/properties [...]

There is, however, an alternative way to satisfy (FC): instead of providing unrestricted sameness conditions, we could have *restricted* conditions that only apply *intrapersonally* and *synchronously*. That is, we could satisfy (FC) through a weaker condition of the following form:

(K1_c) A concept C is the same concept as $C_x \rightarrow$ concept C has property/properties [...] **or it is had by a different subject from C_x**

(As usual, “different subject” also covers different time-slices of the same subject). The second, weaker version of our condition will still satisfy (FC) if the first one does: by hypothesis, C_y belongs to the *same* subject as C_x , so C_y won’t be the same concept as C_x if it doesn’t have the property/properties in brackets.

While both of the above conditions satisfy (FC), however, a restricted condition has an important advantage: it does *not* entail that C_x isn’t shared by other subjects! For consider some subject (other than the one involved in the Frege case) whose concepts don’t have the property in brackets. Clearly, all his concepts *meet* the restricted condition, since they all belong to a different subject from C_x . So restricted conditions are guaranteed to be compatible with the publicity of concepts, since they just don’t apply at the interpersonal/diachronic level.

An example will be helpful. Consider again the localist condition:

(K1_{Edmund}) A concept C is the same concept as SUPERMAN₁ → C has same inferential role as SUPERMAN₁

(where, as usual, the inferential role of SUPERMAN₁ includes [CAPTURED THE ROBBER]). As it stands, this condition entails that SUPERMAN₁ is not shared by someone like Verna, who hasn’t heard the news about Superman. But suppose we *restrict* the condition as follows:

(K1_{Edmund}) A concept C is the same concept as SUPERMAN₁ → C has same inferential role as SUPERMAN₁ **or it is had by a different subject from SUPERMAN₁**

Obviously, all of Verna’s concepts meet this condition, since they satisfy the second disjunct. Therefore, Verna could still have Edmund’s SUPERMAN₁, even though she hasn’t heard the news and doesn’t know about the robber’s capture! At the same time,

the condition will still satisfy (FC), since SUPERMAN₂ does belong to the same subject (Edmund) as SUPERMAN₁.

Restricted conditions seem to have all the advantages of unrestricted ones, without any of the costs: they are just as good at satisfying (FC), but they are guaranteed to be consistent with publicity. Nor is the solution *ad hoc* and unmotivated. By its very nature, (PUB) only applies to concepts had by *different subjects*: communication, agreement/disagreement and intentional generalizations are essentially *interpersonal* phenomena. Conversely, (FC) only applies to concepts had by the *same subject* at the *same time*: the puzzle of rationality in Frege cases is essentially *intrapersonal* and *synchronic*. It's not puzzling that two thinkers can respectively ascribe F and not-F to the same x ; nor is it puzzling that I can ascribe F to x at t_1 and then, having changed my mind, ascribe not-F to the same x at t_2 ! Since (PUB) only applies to concepts had by different subjects and (FC) only applies to concepts had by the same subject, we might have good reasons to provide different sets of sameness conditions corresponding to our two constraints: the first set will apply interpersonally and diachronically, the second intrapersonally and synchronically.

The possibility of a restricted individuation criterion should certainly have received more attention in the literature²⁴¹: as I previously noted, all the criteria reviewed in chapter 4 were *unrestricted*. Nevertheless, I think our first solution is preferable: the idea that a concept can have restricted sameness conditions appears metaphysically dubious. The worry will be declined differently depending on the ontology we are targeting. A Fregean must hold that a concept can have *restricted identity conditions*: Edmund's SUPERMAN₁ is numerically identical to one of Verna's concepts, but *not* to SUPERMAN₂, even though neither of these concepts has the same inferential role as SUPERMAN₁. Does it make sense for a condition on numerical identity with an object x (in this case, the concept SUPERMAN₁) to be restricted in this way? If having a certain inferential role is necessary for SUPERMAN₂ to be numerically identical to SUPERMAN₁, then it seems that condition should also apply to Verna's concepts.

A parallel worry arises for the non-Fregean, who must grant that there are *restricted conditions* on being *type-identical* with SUPERMAN₁. More precisely: the non-Fregean must hold that there is a type T, such that Verna's concepts can belong to T even if they don't have SUPERMAN₁'s role, while Edmund's concepts cannot. Being connected to [CAPTURED THE ROBBER] is thus necessary for T-membership, but *only* for concepts had by Edmund at the same time as SUPERMAN₁. Again, allowing for such restricted typing conditions might seem metaphysically dubious, although admittedly the worry appears more serious for the Fregean. After all, the non-Fregean could respond that we can type concepts any way we like: we can simply stipulate that some typing conditions for T only apply intrapersonally/synchronously, but not interpersonally/diachronically.

²⁴¹ But v. the discussion of intrapersonal/interpersonal typing criteria in Schneider (2011, pp. 126-132); v. also Block (1995, pp. 185-187).

Since the non-Fregean reply seems more promising, the appeal to restricted sameness conditions might require adopting a non-Fregean ontology for concepts. Given that our first solution does not have potentially troublesome metaphysical consequences and can be accepted on both of our ontologies, I believe it is preferable on balance. Fortunately, we need not decide the issue here: my goal in the next section is to convince you that holism should have a place in your individuation criterion for concepts, and my arguments for this claim will go through no matter which of the two solutions you prefer.

4.2. A Place For Holism

As you will recall, in chapter 3 I defended a version of the publicity principle for intentional generalizations and noted how the original Block-Schneider version of holism would be inconsistent with that principle. Then (sect. 4.1), I suggested weakening holism so as to make it compatible with publicity: the holist could abandon the claim that *all* concepts are individuated by their global inferential role, holding instead that only *some* of them are. This gave rise to a worry: maybe we can make a holistic criterion compatible with publicity, but that doesn't show we also have positive reasons to accept such a criterion. Why think that *any* concepts are holistically individuated? And, more generally, why should the modified holistic picture I sketched "have a place" in our best theory of concepts?

In this section, I will argue for the inclusion of a modified holistic condition in our individuation criterion for concepts. First, I will try to convince you that the condition in question would in fact be compatible with publicity; then, that we also have strong positive reasons to accept such a condition. I will then conclude with some general remarks about the role of holism in a theory of concepts.

Holism and other criteria

The first thing to note is that, even in the original Block-Schneider version, holism does no worse than many of its competitors: as noted in chapter 4, several other individuation criteria that satisfy (FC) will also end up violating (PUB). In particular, Fodor's "Conceptual Atomism" seems to present the very same problem as holism²⁴². If concepts are individuated by their reference *plus* their syntactic properties *qua* LoT symbols, no two subjects will have any concepts in common: on both "physicalist" and "functionalist" accounts of LoT syntax, token symbols belonging to different systems will almost always differ in their syntactic properties. Ironically enough, then, Fodor's atomism falls prey of his own arguments against holism, and so do the other views we

²⁴² V. condition (d) in ch. 4 (sect. 4.2) for discussion.

examined (v. e.g. ST's originalism). In conclusion, holism is no worse than other unrestricted criteria when it comes to dealing with publicity.

A problem for everyone

Not only is holism on the same boat as other theories of concepts; more importantly, it can be made *compatible* with the publicity principle by employing one of the two solutions described in the last section. Holists and non-holists alike face the following problem: finding an individuation criterion that can satisfy (FC) while being consistent with (PUB). I suggested two ways to develop such a criterion:

First Solution (Living With Unshared Concepts)

- To satisfy (FC), provide *unrestricted* conditions for concept sameness.
- These conditions will entail, for many concepts involved in contradictory beliefs, that they are *not shared* (cf. Verna and SUPERMAN₁).
- But this is *compatible* with the publicity principle, since many *other* concepts (e.g. SUPERMAN) are still widely shared.

Second Solution (Restricted Individuation Conditions)

- To satisfy (FC), provide *restricted* conditions for concept sameness.
- Since the conditions are restricted, they will *not* entail that concepts involved in contradictory beliefs are not shared.
- Since they are just as *fine-grained* as unrestricted conditions, however, they will still entail that such concepts (e.g. SUPERMAN₁-SUPERMAN₂) are distinct, thus satisfying (FC).

Crucially, everyone will have to employ one of these solutions in his individuation criterion: everyone must grant *either* that many concepts are not shared, *or* that they are shared but have fine-grained, restricted sameness conditions. Once we apply one of the above solutions to the original Block-Schneider account, then, we will get a version of holism that's guaranteed to be consistent with (PUB), as the anti-holist himself must recognize²⁴³!

²⁴³ I will be brief in my treatment of this point, as I have already given details about how to implement each solution in the previous section.

If we decide to apply the first solution, the resulting holistic view will go as follows. We will abandon the original Block-Schneider thesis that *all* concepts are individuated by their global inferential role. Some concepts will still be individuated in this way, most notably those concepts that are involved in the contradictory beliefs of rational subjects. For instance, Edmund's SUPERMAN₁ will be individuated by:

(K1_{Edmund}) A concept C is the same concept as SUPERMAN₁ → C has same global inferential role as SUPERMAN₁

It follows that any concept which doesn't have that role is different from SUPERMAN₁, which is therefore not shared by other intentional subjects (e.g. Verna). This, however, does not constitute a violation of publicity, for other concepts are individuated in a non-holistic way. For instance, there might be a concept SUPERMAN which is not individuated by its global inferential role and which Edmund and Verna do share. (That concept might e.g. be individuated by the standard Superman-ish role). Bottom line: concepts involved in Frege cases are holistically individuated and not shared (to satisfy (FC)), while other concepts are non-holistically individuated and shared (so (PUB) is respected)²⁴⁴. (As you might have noted, this is just the modified holistic theory sketched at the end of chapter 3).

If we apply the restriction solution, the original Block-Schneider picture will be modified in a different way. We will now drop the claim that, for *any two* concepts C₁-C₂, C₂ is the same as C₁ only if it has C₁'s global role; we will instead take this condition to apply only if C₁-C₂ belong to the same subject. If C₁-C₂ are had by different subjects, C₂ can be the same as C₁ even if the two concepts *differ* in some of their connections. Edmund's SUPERMAN₁ will then be individuated by the restricted holistic condition:

(K1_{Edmund}) A concept C is the same concept as SUPERMAN₁ → C has same global inferential role as SUPERMAN₁ ***or it is had by a different subject from SUPERMAN₁***

As noted in the previous section, this condition will be just as effective as an unrestricted one in satisfying (FC), but it will not entail that SUPERMAN₁ is not shared by other subjects. Once we modify the original Block-Schneider account in this way, then, we can take *all* concepts to be individuated by holistic conditions, as long as the conditions in question only apply intrapersonally and synchronically.

I have described two ways to weaken the Block-Schneider account in order to make it compatible with publicity. Again, notice that the anti-holist himself will have to apply one of these strategies to his own individuation criterion in order to make it consistent with publicity. Therefore, he'll have to recognize that, once the holist applies one of the two strategies to his account, the result will also be compatible with (PUB). In

²⁴⁴ V. sect. 4.1 *supra* and ch. 3 (sect. 4.1) for details about how to decline this solution on Fregean/non-Fregean ontologies.

conclusion, everyone must grant that a suitably modified holistic view will not violate publicity after all.

Satisfying (FC)

We have shown that a holistic criterion is no worse than many of its competitors, and (more importantly) that we can modify it to make it consistent with (PUB). But what positive reasons can be offered for accepting such a criterion?

A first reason is that holism will do a great job at satisfying (FC). Subjects involved in Frege cases typically have pairs of concepts that *differ* in some of their inferential connections: cf. Edmund, Lois, Peter and so forth. If having different connections is enough for the concepts in question to be distinct, as the holist maintains, then (FC) will be satisfied in all these cases.

Indeed, we can show in a principled way that holism will satisfy (FC) in *all* the cases covered by the principle. Suppose for *reductio* that holism does not satisfy (FC) in a case *c*. Then there are two concepts C_1 - C_2 such that:

- 1) C_1 - C_2 are involved in the contradictory beliefs of a rational subject *S* at time *t* about some object *x*.
- 2) C_1 - C_2 stand in exactly the *same* inferential connections for *S* at *t*.

This would be the only kind of case in which (FC) requires C_1 - C_2 to be distinct (since they are involved in contradictory beliefs) but a holistic account does not entail that they are (since they have the *same* connections). The case, however, is not only hard to imagine; it is downright impossible, as the following argument shows:

- a) C_1 - C_2 are involved in contradictory beliefs (by (1)).
- b) So C_1 is inferentially connected to [IS F], and C_2 is inferentially connected to [IS NOT F] (from (a)).

(By definition, if the two concepts are “involved” in contradictory beliefs, the subject accepts the thoughts C_1 IS F and C_2 IS NOT F. So he will also be disposed to infer from C_1 to [IS F] and from C_2 to [IS NOT F].)

- c) C_1 - C_2 have the same inferential connections (by (2)).
- d) But then C_1 is inferentially connected to [IS NOT F], and C_2 is inferentially connected to [IS F] (from (b), (c)).

- e) So C_1 is inferentially connected to *both* [IS F] and [IS NOT F], and C_2 is inferentially connected to *both* [IS NOT F] and [IS F] (from (b), (d)).
- f) But if C_1 - C_2 have such inferential connections, our subject will arguably be *irrational, contra* assumption (1).

Our subject would be accepting *both* C_1 IS F and its negation C_1 IS NOT F; similarly, he would be accepting *both* C_2 IS F and its negation C_2 IS NOT F. But, according to the principles of rationality presupposed by our account of Frege cases, this would be an irrational doxastic behavior (cf. the “Standard Explanation” in ch. 4, sect. 2.2). Conclusion: there are no two concepts of which assumptions (1)-(2) are both true: holism satisfies (FC) in all the cases covered by the principle.

The best account of Frege cases?

Maybe holism is good enough to satisfy (FC), but why think it will be superior to the other candidate conditions examined in chapter 4? After all, those conditions can also be patched by employing one of the two strategies considered above. So they can also be made compatible with (PUB), and they might be just as effective as holism in satisfying (FC).

I won’t be able to make a detailed comparison between holism and every other condition examined in chapter 4, so I will limit myself to some general remarks. We already know that some of the conditions discussed in chapter 4 fail to satisfy (FC) in at least some Frege cases. For instance, ST’s originalist view cannot account for Paderewski-like cases: Peter has contradictory *de re* beliefs about Paderewski, but the concepts involved in those beliefs are *not* distinct, since they have exactly the same origin (they are associated with the same public name). Clearly, then, holism will be superior to originalism, and to all the other accounts that make mistaken predictions about some of the relevant Frege cases.

More generally, holism presents a clear advantage over its rivals. While we have a principled argument showing that a holistic condition will satisfy (FC) in all the relevant cases (cf. *supra*), no such argument is available for other conditions, such as physicalist accounts of LoT symbols (v. condition (d) in ch. 4)²⁴⁵. So we have no guarantee that a physicalist account will make the right predictions about all the cases covered by (FC); *ceteris paribus*, this gives us good reasons to go for holism.

What about the other condition examined in chapter 4, i.e. the modified localist account sketched in section 4.3²⁴⁶? On that view, a concept’s inferential role does not include *all* of its inferential connections; however, some concepts have a more fine-grained role than others. For example, [CAPTURED THE ROBBER] will be part of

²⁴⁵ Sect. 4.2. Recall that the other possible version of (d), a *functionalist* one, simply collapses into condition (c).

²⁴⁶ Recall that the unmodified localist account could not satisfy (FC): v. condition (c), ch. 4 (sect. 4.2).

SUPERMAN_1 's inferential role, but not of SUPERMAN 's role. So the account is consistent with (PUB) (SUPERMAN is still shared by Edmund and Verna) and it satisfies (FC) (Edmund's SUPERMAN_1 is still distinct from his SUPERMAN_2). Moreover, a similar strategy can presumably be extended to other Frege cases. Why should we prefer a modified holistic condition to a modified localist one?

The problem with the localist account is that its notion of inferential role is too flexible, so flexible that it collapses into holism! The localist, recall, endorses the following two theses:

Localist definition of inferential roles: the inferential role of a concept C includes *some* (but not *all*) the inferential connections in which C stands at a time t for a subject S

Necessary condition (c): C_2 is the same concept as $C_1 \rightarrow C_2$ has same inferential role as C_1

Now consider an arbitrary Frege case involving two concepts C_X-C_Y . The localist will have to find some inferential connection I_X such that C_X , but not C_Y , has I_X : he will then take I_X to be part of C_X 's role, so that the two concepts come out different. At the same time, however, not *all* of C_X-C_Y 's connections can be part of their roles, or the account would be holistic. For instance: in Edmund's case, [CAPTURED THE ROBBER] will be part of SUPERMAN_1 's role, while other inferences (e.g. [IS HEROIC]) will be among the concept's connections without being included in its role.

We can now see why a modified localist picture would simply collapse into holism. In many Frege cases, the connection I_X which distinguishes C_X from C_Y might not be "special" in any way: it might be just like all the other connections in which C_X stands, in which case there won't be any principled reason to include I_X in the concept's role while leaving the other connections out. This is precisely the case with Edmund. His disposition to go from SUPERMAN_1 to CAPTURED THE ROBBER is one he has only recently acquired, which is no way a "core" component of his concept of Superman, and which he would quickly drop if his evidential situation were to change. Moreover, it is a very "specific" inferential disposition, one based on information that other Superman-thinkers might well fail to have (v. Verna). Why, then, would *that* inference be included in the concept's role, and the inference from e.g. SUPERMAN_1 to HEROIC be left out? Why not include the latter and leave the former out, for instance? The localist faces a version of the notorious "principled basis" worry²⁴⁷: his account seems entirely arbitrary. To avoid the charge of arbitrariness, the only option would be to embrace holism and simply include *all* of SUPERMAN_1 's connections in its role.

The worry is made even deeper by noting that the localist has *already conceded* that concepts involved in contradictory beliefs (e.g. Edmund's SUPERMAN_1) are often *not*

²⁴⁷ V. Fodor (1987, ch. 3), and Fodor and Lepore (2002). We will go back to this in ch. 6 *infra*.

shared by subjects covered by (PUB) (e.g. Verna). But then why should we only include *some* of their connections in their role, if this will not guarantee their shareability? The localist commits to an unprincipled distinction between individuating/non-individuating inferences without gaining any advantages over the holist. I conclude that, when compared with modified localism and the other sameness conditions examined in chapter 4, a suitably modified holistic condition will still be the best way to satisfy (FC).

The demands of (FC)

There is one final consideration favoring the inclusion of a holistic condition in our individuation criterion for concepts. In section 2.1, I showed (FC) to entail the following sameness condition:

(C) (two concepts C_1-C_2 are had by the same subject S at the same time t and they are the same concept) \rightarrow (they do not have contradictory inferential connections)

I then went on to show how *fine-grained* this condition is (section 2.2): the presence of *any two* contradictory connections will be sufficient for two concepts had by the same subject to be distinct, no matter how many other connections they might have in common. To illustrate the point further, I also noted how the *unrestricted* version (C^*) of condition (C) would be straightforwardly *inconsistent* with publicity, since it would individuate concepts had by different subjects too finely.

Since (FC) entails (C), we know our individuation criterion will have to include a fine-grained condition *anyway*: that condition being (C) itself, of course. In fact, notice that (C) is very close to being *holistic*. Like a holistic condition, it is extremely fine-grained, and it would be incompatible with publicity if applied unrestrictedly to all concepts. It is also an “inferentialist” condition, since it takes the presence/absence of certain inferential connections to be relevant for a concept’s identity. The only difference between holism and (C) is that, according to holism, two concepts are different whenever they have *different* connections. Condition (C), on the other hand, only claims that two concepts are different whenever they have *contradictory* connections: if C_1 is connected to [IS F] and C_2 is connected to *neither* [IS F] *nor* [IS NOT F]²⁴⁸, it doesn’t follow from (C) that C_1 is different from C_2 ²⁴⁹.

²⁴⁸ (In this case, the subject is simply agnostic about whether C_2 is F or NOT F).

²⁴⁹ In fact, I believe a plausibly stronger version of (FC) might well entail a holistic condition for concept sameness. In particular, I suspect a holistic condition will be required if we want to give an account of certain further problems arising in Frege cases, especially puzzles of behavior and cognitive significance; I plan to explore the issue in future work. V. Block (1995) and Schneider (2009a, pp. 531-535; 2011, pp. 104-109 and 118-120), who both offer principled arguments for holism along these lines. (Schneider’s argument partially inspired the arguments in sects. 2.1-2.2, although I’m not sure I understand all its details).

While strictly speaking weaker, (C) is analogous to a holistic condition in all the relevant respects: it would be incompatible with (PUB) if applied unrestrictedly, it is very fine-grained, and it has an “inferentialist” component. This has important consequences, since I have shown (C) to be *directly entailed* by (FC): anyone who accepts the Fregean constraint (not only those with holistic sympathies!) will have to include (C) in his individuation criterion. On everyone’s account, then, our best theory of concepts will comprise a condition that is quasi-holistic. In light of this, including a fully holistic condition should seem more appealing: after all, I have already offered several reasons to do so, and we are already committed to a fine-grained inferentialist condition like (C) in any case.

I have argued that holism should “have a place” in our best theory of concepts. More specifically: depending on which of our two solutions we adopt, an individuation criterion should either include *unrestricted* holistic conditions for *some* concepts, or *restricted* holistic conditions for *all* concepts. I will conclude with two brief remarks. First: once we realize that the Block-Schneider account can be modified in one of these two ways, the debate over holism “deflates” as insubstantial. On both of our solutions, holism will be *compatible* with other theories of concepts. On our first solution, we can still individuate several concepts in a non-holistic fashion (e.g. by their reference, or by their local inferential role). On our second solution, we can provide non-holistic sameness conditions for concepts had by different subjects, while having holistic conditions for concepts had by the same subject, at the same time. The two sets of conditions will be entirely compatible: being restricted, the first set will only apply interpersonally/diachronically, while the second set will only apply intrapersonally/synchronously. Once we modify our holistic account in one of these two ways, every appearance of conflict with non-holistic views disappears: both of them can be incorporated in our best theory of concepts (more on this soon).

Second remark: if concepts are to play the “mode of presentation” role, then they must be individuated holistically²⁵⁰. According to a venerable tradition in philosophy of mind, the rationality of subjects involved in Frege cases is explained by their having *distinct modes of presentation* for the same object x ²⁵¹. Once we also identify modes of presentation with concepts, it follows that every subject involved in a Frege case must have distinct concepts for the same x : indeed, this is what (FC) amounts to. To build a theory of concept individuation that can account for this, I have suggested individuating modes of presentation/concepts by their global inferential role; this “cuts” concepts finely enough to predict, for every subject involved in a Frege case, that he does have distinct concepts for the same object. Conclusion: if concepts are to “play the mode of presentation role”²⁵² and account for certain facts about rationality, then they should be

²⁵⁰ Again, v. Schneider (2009a, 2011), who arrives at the same conclusion through a different route.

²⁵¹ Ch. 4, sect. 2.

²⁵² The expression is in Schiffer (1990).

individuated holistically. Our venerable tradition is committed to a form of holism about concepts.

4.3. A Picture Of Concepts

I have argued that holism should have a place in our best theory of concepts; this was one of my two goals in chapters 4-5. My other goal was to provide independent support for the contextualist version of publicity developed in chapter 3: more precisely, to convince you that, if intentional generalizations do require concepts to be shared by intentional subjects, such requirements will differ depending on the context of the generalization, and speakers' intentions will play a crucial role in determining which concepts have to be shared in a given context. In this final section, I will show how this picture of publicity becomes extremely plausible on at least one of the two solutions I proposed in section 4.1. I will then conclude by sketching my own picture of concept individuation and possession.

Throughout chapters 4-5 I have been assuming a generic, “non-contextualist” version of the publicity principle: I have been presupposing that intentional generalizations will require the subjects they cover to have certain specific concepts in common, but not that contextual features will play any role in determining what concepts must be shared²⁵³. This is because I wanted to convince you of the following: *if you find the principle of publicity attractive, then you should endorse a contextualist version of it.* In light of our discussion of (FC) and (PUB), I hope you will now find this claim plausible; as I will now argue, on at least one of our two solutions (the most plausible one indeed) we are almost forced to give a contextualist reading of (PUB).

Recall what our first solution (sect. 4.1) amounted to: (PUB) only requires *some* concepts (those involved in intentional generalizations) to be shared, but this is compatible with *other* concepts (those involved in contradictory beliefs) *not* being shared. For instance, this is what happens with Edmund and Verna, who share SUPERMAN but not SUPERMAN₁-SUPERMAN₂. But now the question arises: *what* concepts must be shared by the subjects to which a generalization applies? If *unshared* concepts like SUPERMAN₁ are involved in our intentional generalizations, then such generalizations will not be interpersonally applicable! What guarantees that this won't happen?

Our contextualist picture provides a convincing answer to this question. (This point was discussed extensively in chapter 3 (sect. 4.1), so my discussion will be brief). In short: speakers will generally *not* make reference to holistically individuated, unshared concepts in their normal uses of intentional generalizations. Suppose an ordinary speaker utters:

²⁵³ Here I will ignore other aspects of the publicity principle, having to do with communication, agreement/disagreement etc..., to focus on generalizations alone.

G5) If a subject S believes that Clark can fly and believes that Clark just jumped off a skyscraper, then other things being equal S will not be worried.

On our first solution, there are several concepts of Superman/Clark that our speaker might be “providing” in his utterance of (G5). For instance, he might be referring to a concept with the same global inferential role as Edmund’s SUPERMAN₁, in which case (G5) will only apply to Edmund (he is the only one to have that concept). Clearly, however, there is no reason for ordinary speakers to refer to such ultra-specific concepts in their ordinary folk-psychological practices. Much more plausibly, our speaker is referring to a more coarsely individuated concept SUPERMAN; he might for instance intend the generalization to apply to anyone who has a concept with the standard Clarkish role, or indeed to anyone who has a concept referring to Clark at all. In both cases, someone will satisfy (G5)’s antecedent as long as he believes <Superman, can fly> under the coarsely individuated SUPERMAN. Clearly, then, (G5) will be widely applicable, since several subjects other than Edmund have the concept in question.

In sum: our solution does presuppose that several concepts are holistically individuated, but (thanks to contextualism) these concepts will *not* be the ones involved in our intentional generalizations. Ordinary speakers have no reason to refer to such concepts when uttering a generalization like (G5); they will rather refer to widely shared concepts like SUPERMAN. Assuming a contextualist version of publicity, then, speakers’ intentions will ensure that only non-holistic concepts are required in order for someone to fall under a generalization like (G5).

Now, in *some* contexts a speaker might decide to refer to holistically individuated concepts: consider for instance a sophisticated theoretical context in which we explicitly postulate the existence of such concepts to answer certain theoretical needs, such as accounting for (FC). Someone might then decide to refer to such concepts when uttering an intentional generalization; he might e.g. decide to refer specifically to Edmund’s SUPERMAN₁, in which case the generalization will not apply to anyone but Edmund. Notice, however, that if a generalization was uttered by a speaker with this kind of intention, it would be intuitively *correct* to treat it as non-interpersonally applicable! If someone (say, Edmund himself) intends to refer to Edmund’s “holistic” concept of Superman/Clark when uttering (G5), why should we take the antecedent of (G5) to be satisfied by anyone but him? Contexts in which a speaker refers to holistically individuated concepts pose no threat, since it would be mistaken to treat the corresponding generalizations as interpersonally applicable.

In conclusion: I have shown that, once we assume our first solution to the tension between (FC) and (PUB), we will also have strong reasons to endorse a contextualist version of the publicity principle. Since the alternative solution, which appeals to restricted conditions, has problematic metaphysical consequences, this gives us good reasons to go for contextualist publicity. (Also, notice that, while the restriction solution might not *motivate* contextualist publicity, it is certainly *consistent* with it). Bottom line:

if you like the publicity principle, you should like the contextualist version of it that was developed in chapter 3.

My general aim in chapters 4-5 was to provide further support for two views that were developed in chapter 3: a modified version of holism and a contextualist model of publicity. I hope my discussion of the relationship between (FC) and (PUB) has made these views look more attractive to the reader. To conclude our discussion, I would now like to integrate those two views in a more general picture of concept individuation and possession. This is the picture that emerges from my arguments in chapters 1-5; I call it a “picture” since it is partly a substantive theory about the individuation/possession conditions of certain concepts, and partly a framework within which one can develop such a theory. In lack of a better term, I will call it a “*pluralist-contextualist*” picture of concepts. (A “*pluralist-localist*” theory has recently been defended by Daniel Weiskopf (2009 a,b), from whom I borrow the term “*pluralism*”. I am much indebted to some of Weiskopf’s insights, although my view rejects his localism and is a lot more radical: v. ch. 6 (sect. 2.3) for a comparison).

We started our discussion with two questions: what are the *individuation conditions* of concepts? And what are their *possession conditions*? On my picture, there is no single answer to these questions: concepts are individuated in a variety of ways. In particular, *some* concepts should be individuated *holistically*, by their global inferential role, in order to satisfy the Fregean constraint: these, of course, are the concepts involved in the contradictory beliefs of rational subjects, like Edmund’s SUPERMAN₁-SUPERMAN₂. Other concepts, however, must be individuated more coarsely. According to our best theory of attitude ascriptions, we often make reference to other people’s concepts in our reports. If no concepts were shareable, generalizations embedding such reports would not apply to more than one subject; there must be some shared concepts for ordinary speakers to refer to in their intentional generalizations. And, luckily, there are: these are the concepts that are not individuated by their global inferential role, like Edmund and Verna’s SUPERMAN^{254,255}.

How, then, are shared concepts individuated? This is a difficult question, and not one I will be able to answer in full. Again, I don’t think there is a single answer. Some

²⁵⁴ I am presupposing something like our first solution here. However, my pluralist-contextualist picture could also be developed along the lines of our second solution. We could then say that concepts are individuated in *multiple ways*: different sets of sameness conditions apply at the intrapersonal/synchronic and interpersonal/diachronic level. More specifically, the sameness conditions applying intrapersonally/synchronously are holistic, while those applying interpersonally/diachronically are coarser (for instance, they might simply require sameness of local inferential role. V. sect. 4.1 for more details about the appeal to restricted sameness conditions). We could then draw analogous “*pluralist*” conclusions regarding the conditions for concept possession (v. *infra*). While differing in some of their details, our two solutions push towards the same general picture of concepts.

²⁵⁵ As usual, all these claims will be declined differently depending on our background ontology. For reasons of space, I must leave it to the reader to construe a Fregean and a non-Fregean version of the picture (v. sect. 4.1 *supra* for instructions on how to do this).

concepts might be individuated by their local inferential role. When speakers make ascriptions of the form “S believes that Clark can fly” and use generalizations embedding that ascription, they will sometimes refer to a concept CLARK which is individuated by the standard Clark-ish role. But it would be wrong to assume that this will happen in all, or even most contexts²⁵⁶. Some thinkers might not have a concept with the standard Clark-ish role; for instance, they might ignore that the shy reporter with glasses is called “Clark”. Yet, a speaker could intend his generalizations to cover these “abnormal” thinkers as well as more ordinary ones. In these contexts, the speaker might simply refer to a concept (call it CLARK*) which is individuated by *part* of the standard Clark-ish role (e.g. only by [SHY REPORTER WITH GLASSES]): his generalizations will then apply more widely, covering abnormal thinkers as well as ordinary ones. In other contexts, a speaker might even intend his generalizations to apply to anyone who has *some* concept referring to Superman/Clark, no matter what his inferential dispositions are. In these contexts, our speaker will make reference to a concept (call it CLARK_{REF}) which is purely individuated by its reference, i.e. by its referring to Superman/Clark: here, any subject who is capable of having thoughts about Superman/Clark will fall under the relevant generalizations, no matter how “deviant” his inferential dispositions towards Superman/Clark might be²⁵⁷.

On our picture, then, there will be a multitude of concepts “in the vicinity” of the concept CLARK; while that concept might be individuated by the local Clark-ish role, there will also be concepts (CLARK*, CLARK**...) which are individuated by parts of that role, and even concepts (CLARK_{REF}) which are simply individuated by their reference and thus shared even more widely. But the heterogeneity of concepts does not stop here. On many views, concepts are at least *partially* individuated by their reference: two concepts C₁-C₂ cannot be the same concept unless they refer to the same object/property²⁵⁸. Unfortunately, this widely endorsed position makes it impossible to account for the shareability of *indexical* concepts such as the first-person concept: clearly, my I-concept and your I-concept refer to different individuals. And yet, the same arguments that move us to recognize the publicity of non-indexical concepts also apply to indexical ones. Consider the generalization:

If a subject S believes that he is going to be attacked by a bear, then other things being equal S will be scared²⁵⁹.

²⁵⁶ This is one of the main differences between my picture and standard localist views: v. ch. 6 for more details.

²⁵⁷ In these contexts, “S believes that Clark can fly” will be true of anyone who believes <Superman, can fly> under some concept or other: the truth conditions of attitude reports and intentional generalizations will then be extensionally equivalent to those provided by the Millian.

²⁵⁸ Fodor’s account is a case in point: v. ch. 4 (sects. 3.1-3.2).

²⁵⁹ V. Perry (2000) for the original bear-example.

You and I both satisfy this generalization: under certain circumstances, each of us would have the belief that he was going to be attacked by a bear, feeling scared as a result. In turn, it seems that, when we have the belief in question, we will both think the *same* thought I AM GOING TO BE ATTACKED BY A BEAR. Similar considerations apply to communication: when I say “I am going to be attacked by a bear” you should be able to grasp the same thought I express, even though that thought will pick out a different individual when entertained by you. This is a huge topic²⁶⁰, of course, and one might well resist the idea that we are deploying the same I-concept. I bring this up only to note that a pluralist picture *can* account for the shareability of indexical concepts, by holding that *some* concepts are not individuated by their reference but, say, by their Kaplanian character (which is of course the same for our respective I-concepts).

A pluralist picture of concept individuation brings with it a pluralist picture of concept possession: there is just no single answer to the question “What does it take for a thinker to have a concept”? Some concepts have more demanding possession conditions, while others have less demanding ones. To have Edmund’s SUPERMAN₁, Verna would not only need to share the information Edmund got from the news, but his global set of beliefs about the world; to have his SUPERMAN, she only needs to know about a few notorious features of the famous superhero. If some concepts are individuated by their reference, then their possession only requires the ability to form propositional attitudes about the relevant object/property; *mutatis mutandis* for indexical concepts and other classes of concepts, whose possession conditions might be different yet from those examined so far.

Analogous considerations apply to questions like: “When do two thinkers have the *same concept* of... (some object/property: Superman, time, justice...)?” This question wrongly presupposes that there will be a unique concept for every object/property²⁶¹: on my picture, there is a multitude of concepts individuated in several different ways, and the conditions for two thinkers to have a concept of *x* in common will vary from concept to concept. Edmund and Verna will have the same concept SUPERMAN just in case they share certain ordinary beliefs about Superman; to have the same concept SUPERMAN₁, however, they would have to share Edmund’s global set of beliefs. As we have seen²⁶², context will play a crucial role in making certain concepts more salient than others, thus making it possible for us to say that two thinkers have “the same concept”. For instance, in most contexts we would say that Edmund and Verna have “the same concept” of Superman. However, once someone remarks how Verna hasn’t heard the news and doesn’t share all of Edmund’s information, we will recognize that “in some sense” their concepts differ. This can be straightforwardly captured by our picture: Verna and Edmund do have the same concept SUPERMAN, but they don’t have the same concept SUPERMAN₁. (The phenomenon is of course ubiquitous: the claim that

²⁶⁰ V. Frege (1956), Evans (1982), Kaplan (1989).

²⁶¹ V. Weiskopf (2009b).

²⁶² Cf. ch. 3 (sect. 4.1).

two people have “the same car” will sound true in context c_1 (they have cars of the same type) but false in context c_2 (someone remarked that the two cars differ in color)).

In light of these considerations, the project of providing a *unique* set of sameness conditions for concepts seems misguided. In particular, the attempt of Fodor, Schneider and others to develop a *single* typing criterion for LoT symbols appears doomed to fail. Concepts should be typed in a variety of different ways: typing by global inferential role is useful for certain theoretical purposes (satisfying (FC)) but not others (respecting (PUB)), for which a coarser typing criterion will be appropriate. It’s only by helping ourselves to a *plurality* of typing criteria that we can fulfill the conflicting demands imposed on our theory of concepts by constraints like publicity and (FC).

Chapter 6

Two Dilemmas For Localism

1. Introduction

My goal in this chapter is to discuss an important family of views about concept individuation and possession. This approach to concepts, which I have called “localism”²⁶³, has already been examined in the context of our broader discussion (v. especially chs. 4-5). My goal here will be to analyze in more detail some of the main localist theories on the market, to compare them to the picture of concepts that was sketched in chapter 5, and to highlight those aspects of localism that I find most problematic.

Here is how I propose to proceed. Having summarized the basic principles of localism, I will raise a first dilemma for localist views (sect. 2): the localist must decide whether to count *all* the structured mental representations that psychologists call “concepts” as properly individuating concepts “in the philosophical sense of the word”, or hold instead that only *some* of those representations actually individuate concepts and are required for their possession. I will then examine a view that chooses the first horn, the “pluralist-localist” theory recently defended by Daniel Weiskopf. I will highlight the main points of contact between Weiskopf’s pluralism and my own pluralist picture, as well as their main differences. As we’ll see, pluralist accounts have a great virtue: they can incorporate some important recent findings from cognitive psychology in their theory of concept individuation/possession. At the same time, however, Weiskopf takes psychological theories to provide correct individuation conditions for *all* concepts; I will suggest that this results in an inadequate theory of concept individuation/possession and offer reasons to prefer my own model, on which we only use psychology as our guide to the individuation conditions of *some* concepts.

In section 3, I will move on to discuss those localist views on which the only inferential dispositions that individuate concepts are those *fundamental* dispositions that we would manifest under certain ideal conditions. Such views have to face a second dilemma: they must decide whether *incorrect* dispositions are also allowed in the individuating set. Having noted a few reasons why localist views generally answer in the negative, I will examine the consequences of choosing this horn of the dilemma. In particular, I’ll note that the localist faces a problem of *psychological plausibility*: if concepts are only individuated by correct dispositions, it becomes hard to account for the fact that ordinary thinkers actually *have* concepts, since the “folk theories” that those thinkers accept are often *incorrect*!

In reply, many localists will hold that folk theories *fix the reference* of our concepts and are thus guaranteed to be correct. In section 4, I will examine the problems with this response by focusing on a specific localist view, the “descriptivist” theory of concepts recently defended by Frank Jackson. Once we take reference to be fixed by our inferential dispositions, it becomes hard to account for those “deviant” subjects whose

²⁶³ V. Devitt (1996), Weiskopf (2009a). The view is sometimes referred to as “molecularism” (Fodor and Lepore 2002).

inferential dispositions are radically different from ours; Jackson's view predicts that these subjects' concepts have *different reference* from ours, a prediction that will be plainly unacceptable in number of important cases. Having considered and rejected some possible replies that could be offered in Jackson's defense, I will conclude my discussion of localism with some more general considerations about the conditions for concept individuation and possession.

2. The Dilemma From Cognitive Psychology

2.1. Localism And Its Virtues

As you will recall²⁶⁴, I defined localism as the conjunction of²⁶⁵:

Concept individuation/possession (IRS): a concept C is partially individuated by its inferential role; therefore, a subject S will have C only if he has some mental representation with the same inferential role as C ²⁶⁶.

Localist definition of inferential roles: the inferential role of a concept C is the set of *some* (but not *all*) the inferential connections in which C stands at a time t for a subject S .

To illustrate, consider our usual example: a localist might plausibly take the inference X IS A BACHELOR \leftrightarrow X IS AN UNMARRIED MAN to individuate BACHELOR, and a disposition to perform that inference to be necessary for its possession. So, among all the inferences that a possessor of BACHELOR might be disposed to draw, some ($BACHELOR \leftrightarrow UNMARRIED\ MAN$) will individuate the concept and will be required for its possession, while others (e.g. $BACHELOR \rightarrow LONELY$) will not.

Following common usage²⁶⁷, we can also say that, on a localist view, the concept BACHELOR is a “structured” entity which is “constituted” by the concepts UNMARRIED and MAN. This is a useful metaphor, but it must not be misinterpreted: the intended reading of “inferential structure” is a very weak one. The localist need not claim that concepts are *syntactically* structured (e.g. that BACHELOR is a complex LoT symbol constituted by the symbols UNMARRIED and MAN). He also need not claim that they are *semantically* structured (e.g. that UNMARRIED MAN is the content of BACHELOR). In what follows, the claim that concepts have “local inferential structure” will thus simply amount to the claim that a concept is individuated by some of its inferential connections, and that the corresponding inferential dispositions are required for its possession.

²⁶⁴ V. ch. 2 (sect. 2.3) for a definition of IRS, and ch. 4 (sect. 4.2) for a first definition of localism.

²⁶⁵ Localist theories have been defended by: Weiskopf (2009 a,b), Devitt (2006), Pereboom (1995), Rey (1983, 1985, 2005), Peacocke (1992, 1998a, 2007), Jackson (1998). Some important differences between these views will be extensively discussed in what follows.

²⁶⁶ Notice that this claim can be accepted on both Fregean and non-Fregean ontologies: the Fregean will take the mental representation in question to express concept C , while the non-Fregean will take that representation to be concept C itself.

²⁶⁷ Cf. Weiskopf (2009a).

Localism is an appealing position. First, as noted in chapter 4, the view seems to offer a safe middle ground between atomistic and holistic theories of concept individuation. Holistic views like those defended by Block or Schneider seem to slice concepts too finely, thus precluding publicity. On the other hand, atomistic theories like Fodor's seem to slice concepts too coarsely²⁶⁸: if concepts are not individuated by their inferential roles, what distinguishes coreferential concepts like SUPERMAN/CLARK or WATER/H₂O? On a localist view, these concepts will be distinct (their local inferential roles are different) and yet widely shared (local roles only include a small number of inferences).

As anticipated in our general Introduction, localism also seems to provide a solid foundation for the “conceptual analysis” project. If concepts have inferential structure, then our practice of conceptual analysis will consist in discovering what that structure is, for a number of philosophically interesting concepts (TRUTH, KNOWLEDGE, JUSTICE...)²⁶⁹. Localism might also support a more ambitious version of the project. Suppose conceptual structure is “*reference-fixing*”: *x* falls under a concept *C* just in case it falls under the concepts which constitute *C*'s structure. It will then follow that we can derive substantive conclusions about real-world philosophical kinds from our knowledge of conceptual structure. Once we find out that the inferential role of BACHELOR is [UNMARRIED MAN], we can conclude that something is a bachelor just in case it is an unmarried man. This is a conclusion about bachelors, not about our concept of them. It is a substantive claim about the world²⁷⁰.

Finally, localism seems to capture both the “commonsense” and the “scientific” notion of a concept. In everyday discourse, we often talk about the “concept” of (say) *knowledge*, *God*, *state* etc... that a certain person or community has. When we do this, we seem to ascribe a “theory” or “conception” constituted by a small set of beliefs, and this is precisely what the localist identifies with concepts. Even more importantly, localism seems to fit well with the psychologist's usage of the term “concept”. Cognitive scientists are constantly positing structured mental representations to explain a large body of experimental data concerning categorization, induction, concept learning and so on²⁷¹. These representations, which psychologists call “concepts”, have *local structure*: they include *some*, but not *all* the beliefs that a subject has about the relevant category, and they are usually identified with small, well defined sets of specific inferences. For this reason, neither atomists nor holists can count the structured mental representations posited by psychologists as “concepts”. In turn, this has led many localists to claim a further advantage: the view is explanatorily superior to

²⁶⁸ Atomistic theories have also been defended by: Millikan (1998, 2000), Margolis (1999), Edwards (2009, 2010).

²⁶⁹ V. Peacocke (1998a) and Rey (2005); v. also Goldman (1998, 2007) for a different version of the project.

²⁷⁰ V. Jackson (1998) for a version of the ambitious project. We will go back to this in sect. 4 *infra*.

²⁷¹ More on this soon.

atomism and holism, since it can appeal to a concept's local structure to explain experimental data about categorization and other cognitive tasks²⁷².

2.2. What Does Psychology Tell Us About Concepts?

While localism seems to present some advantages, it is also affected by a problem that neither atomists nor holists have to face. Among the many inferential connections in which a given concept will happen to be involved, how do we select the individuating ones? What is our “principled basis”²⁷³ in excluding some inferences (e.g. BACHELOR → SAD) as “collateral” while including some others (e.g. BACHELOR → UNMARRIED) as individuating?

Our considerations in the last section seem to suggest a straightforward response. Unlike atomists and holists, localists can count the locally structured representations posited by psychology as genuine concepts. For any concept C , then, a localist might simply identify the inferences which individuate C and are required for its possession with those that the psychologist will include in his model of C ²⁷⁴. For instance: the set of inferences that individuate BIRD and are required for its possession *just is* that set of inferences which is identified with the concept BIRD by our best psychological theory.

But what picture of concepts would emerge if the localist decided to follow the psychologist in his account of concept individuation/possession? To answer this question, we must briefly look at the models of concepts that have been proposed by recent cognitive science in the attempt to explain data about categorization and other cognitive competences. We can identify four main models²⁷⁵:

Definitions: A concept with definitional structure consists of a set of features which are taken to be individually necessary and jointly sufficient for category membership: an object will fall under the concept just in case it has each feature on the list. The once dominant “classical theory”, which identified concepts with definitions, has today been entirely replaced by other models due to its inadequacy in explaining experimental data.

Prototypes: A concept with prototypical structure consists of a set of features which are taken to be typical of category members. Features are assigned a certain weight depending on their degree of typicality. For instance, a prototype for BIRD will include highly typical features such as FLIES, NESTS IN TREES, LAYS EGGS... : an object x will then be categorized as a bird just in case it “scores” a sufficient degree of similarity to the prototype in virtue of instantiating some combination of those features. For this

²⁷² Weiskopf (2009a), Prinz (2002).

²⁷³ V. Fodor (1987, ch. 3), and Fodor and Lepore (2002). The worry can be traced back to Quine's (1951) attack on the analytic-synthetic distinction. V. ch. 5 (sect. 4.2) for some discussion.

²⁷⁴ The move is suggested by Pereboom (1995) and Weiskopf (2009a).

²⁷⁵ V. Murphy (2002), Machery (2009) for an overview.

reason, prototypes are “similarity-based” structures: they differ sharply from definitions, since none of the features is individually necessary for category membership and any combination of them is enough for x to fall under the concept.

Exemplars: A concept with exemplar structure consists of a set of features which have been stored in long-term memory after encounters with specific members of the category (my dog, my neighbor’s dog...). An object will be categorized as falling under the concept just in case it shares enough features with stored representations of concrete exemplars. Like prototypes, exemplar concepts are “similarity-based”: an item will be categorized as falling under the concept as long as it has a sufficient number of features in common with stored exemplars, so none of the features is individually necessary. Unlike prototypes, however, exemplar concepts do not encode a general representation which is statistically abstracted from instances of the category, but only specific representations of particular category members²⁷⁶.

Mini-theories: Mini-theories encode “theoretical” information about the relations holding between the properties of category members. In particular, *causal* relations are represented, so that greater weight is assigned to “essential” and “deep” properties, on which less essential and more superficial properties causally depend. Unlike exemplars and prototypes, which simply encode sets of features that are “matched” with the target item to determine its similarity score, the kind of information that is embedded in mini-theories can be quite sophisticated and result from complex reasoning about whether a certain property is more or less causally central.

Now, while traditional theories of concepts tended to subsume concepts under a *single* structural model (“concepts are prototypes”; “concepts are mini-theories”), it is more and more recognized that the mental representations we use to categorize, perform inductions, reason deductively and so on can take a variety of formats. A strong *pluralist* trend is emerging in psychology: *each* of the structures described above can be deployed by ordinary thinkers in the appropriate circumstances²⁷⁷. For any given category, one and the same subject will create, store and employ representations of very different kinds: in most cases, a subject will be able to make use of a prototype, a set of exemplars and a mini-theory²⁷⁸. For instance, each of us will have a prototype and a set of exemplars he can use to categorize birds, *as well as* a mini-theory representing those

²⁷⁶ Note that, just like the other psychological models described here, exemplar concepts will *also* have inferential structure. Simplifying: suppose I have only encountered three birds and encoded in long-term memory the following information: BIRD₁ [F, G]; BIRD₂ [H, I]; BIRD₃ [J, K]. Now suppose I am disposed to categorize something as a bird just in case it has any three features in common with my exemplars. We might then represent my exemplar concept BIRD as the following set of inferential dispositions: X IS A BIRD \leftrightarrow X HAS (F, G, H) OR (F, H, J) OR ... (and so on for all the other possible combinations).

²⁷⁷ For a review of the evidence, v. Machery (2009), Weiskopf (2009 a,b).

²⁷⁸ Rips (1989), Piccinini and Scott (2006).

“hidden”, essential properties that causally underlie a bird’s superficial features. Which representation is used will depend, among other things, on the cognitive task at hand, on the nature of the real-world category involved, and on the experimental setting in which the subject operates.

There is also evidence for other kinds of conceptual variation. Within the same experimental setting, different subjects will use representations with different structures: some subjects might opt for exemplars, others for prototypes²⁷⁹. And even when two subjects employ representations of the same kind, there are still going to be interpersonal variations: my prototype or exemplar for dogs might be significantly different from yours depending on the kind of dogs I have encountered²⁸⁰.

The emergence of pluralism creates a first dilemma for localism. Do *all* of the heterogeneous mental representations posited by psychologists individuate corresponding concepts? Each of these mental representations has local inferential structure, and to each of them the psychologist applies the label “concept”. Will the localist acknowledge that the psychologist’s use of the term is correct, and that for each of these locally structured representations there is a corresponding concept which is individuated by that structure? Depending on how they answer this question, two radically different versions of localism can be distinguished; let’s discuss each of them in turn.

2.3. Weiskopf’s Pluralist Localism

When confronted with pluralist data, the localist might decide to take psychological theories at face value and continue to use them as a principled basis to select individuating inferences; this localist will count as concepts *all* the structured mental representations to which our best psychological theories apply the label “concept”. More precisely, let “pluralist localism” (PL) designate any theory of concepts that endorses our basic set of localist assumptions (sect. 2.1) *plus*:

Pluralism: For any structured mental representation that our best psychological theory of concepts calls a “concept”, there is a concept that is individuated by that mental representation.

A sophisticated (PL) view has recently been developed by Daniel Weiskopf (Weiskopf 2009 a,b)²⁸¹; in this section, I will compare Weiskopf’s pluralist localism with the contextualist-pluralist view I developed in chapter 5²⁸². I will first examine

²⁷⁹ Malt (1989), Smith and Minda (1998).

²⁸⁰ Barsalou (1989, 1993).

²⁸¹ Since Weiskopf’s view exemplifies well all the relevant features of (PL), I will use it as my target in what follows.

²⁸² V. sect. 4.3.

some common features of the two views: in particular, both of them have the advantage of being able to accept psychological models of concepts as correct, thus giving an empirically informed account of concept individuation and possession. I will then note two main differences. First, Weiskopf takes *all* concepts to be individuated by the local representations described by cognitive psychology. Going back to some of the arguments from chapter 5, I will note how this makes it hard for a theory of concept individuation to satisfy the Fregean constraint. This gives us reasons to prefer my own pluralist picture, on which *some* concepts are locally individuated in the way suggested by the psychologist, while *others* are individuated differently (e.g. holistically) to satisfy constraints such as (FC). I will then note a second difference between Weiskopf's view and mine, arguing that (for those concepts which are inferentially individuated) we should not take their inferential role to be part of their *content*, but simply an essential component of their individuation conditions.

Let's say a bit more about Weiskopf's (PL). According to Weiskopf, a concept C has two types of content: a *referential content* (RC), which is the category/individual to which C refers, and a *conceptual content* (CC), which is the set of semantic features posited by our best psychological model for that concept. For instance: a good psychological theory will ascribe to me a certain bird-prototype to explain my categorization behavior with respect to birds. This prototype will be (more or less) different from the one you employ. Having prototype structure, it will also be different from my exemplar and mini-theoretic representations for birds. On Weiskopf's view, these structured representations are *all* concepts: I have a prototype-concept BIRD with a different CC from both your BIRD-prototype and my other concepts for birds (exemplars, mini-theories...). There will then be a *plurality* of *locally structured* bird-concepts, each corresponding to one of the bird-representations posited by the psychologist.

Three aspects of Weiskopf's picture will be especially important here:

- 1) Concepts are individuated by *both* their CC and their RC. Crucially, a concept's CC will only incorporate *some* of the information associated with that concept; therefore, a concept is always individuated by *some* but not *all* of its inferential connections²⁸³.
- 2) The inferences which individuate a concept constitute part of its *content* (its conceptual content or CC). For instance, the set of inferential dispositions which constitute my bird-prototype are part of the content of one of my bird-concepts.
- 3) Concepts with *different* CCs will often pick out the *same* category. This is an essential aspect of Weiskopf's pluralism: for one subject to have multiple concepts

²⁸³ V. especially pp. 134-38, where Weiskopf argues that psychological models are local and non-holistic (all quotes are from Weiskopf (2009a)).

of the same category, it must be possible for those concepts to have different CC but same RC. Similarly for concepts had by different subjects. This constraint has a further important consequence: the CC of a concept *cannot fix its reference*. If it did, many concepts which should have the same RC would end up referring to different categories. Suppose I have both a mini-essentialist and a prototype concept for birds. If their CCs were reference-fixing, these two concepts would have different RCs: the first one would pick out a category whose members share the deep essential features of birds, while the second one would pick out a category whose members share the superficial bird-features. The latter (but not the former) *would* include robot-birds that behave and look exactly like prototypical birds, and it would *not* include a penguin that lost its beak and wings. For these two concepts to represent the same category, their RC must be fixed by something *other* than their CC. For this reason, Weiskopf takes reference to be determined by a “purely externalist” mechanism (e.g. a causal chain or a deferential mechanism); indeed, conceptual content is so radically “disconnected” from reference that, in some cases, a concept’s CC will be *entirely false* of its RC (pp. 142-43)!

We now have a better sense of what a localist view will look like if it decides to embrace the first horn of our dilemma. What are the points of contact between (PL) and the pluralist view I sketched in chapter 5? What are the main differences²⁸⁴?

Like Weiskopf, I also take several concepts to be individuated by their inferential connections: v. for instance the holistically individuated concepts described in chapters 4-5, or the locally individuated concept CLARK introduced in chapter 3 during our discussion of contextualism. Even more importantly, I also allow for concepts to be individuated in multiple ways and to have heterogeneous structures. This is an important advantage for pluralist pictures, since it allows them to count those mental representations that psychologists call “concepts” as proper concepts “in the philosophical sense of the word”. When the psychologist claims that prototypes, exemplars and theories are all “concepts”, pluralists like Weiskopf or me need not take the psychologist to be making a *false* claim, nor do we need to interpret him as referring to a *different* kind from the philosopher. As previously noted, this brings further advantages: for a large class of concepts, the pluralist can offer an empirically informed account of their individuation conditions and use that account to explain a wide range of empirical data.

This leads us to a first difference between my view and Weiskopf’s: the problem with (PL) is that it is *not pluralist enough*! While *some* concepts are to be identified with the locally structured representations described by psychology, I have also given reasons to think that not *all* concepts are individuated in this way. In particular, I have argued that some concepts (those involved in contradictory beliefs, such as Edmund’s

²⁸⁴ Weiskopf’s (PL) has been a major source of inspiration for my view. I have also been influenced by Machery (2009) and Prinz (2002), but I cannot discuss their views here.

SUPERMAN₁-SUPERMAN₂) should be individuated by their *global inferential role*: this is what's needed to satisfy (FC) in every possible Frege case²⁸⁵.

Now, as you might recall, one possible alternative to holism was a “modified localist” view²⁸⁶, on which the inferential role of concepts like SUPERMAN₁-SUPERMAN₂ is:

- *Very specific* (it includes inferences like [CAPTURED THE ROBBER]).
- But *not holistic* (it does not include *all* of the concept's connections, e.g. it leaves out [IS HEROIC]).

In reply, I argued that:

- a) The localist cannot include the former inferences in the concept's role without *also* including the latter: there is no “principled basis” he can use to distinguish the former inferences as the only individuating ones.
- b) The view would be unmotivated. Once we include very specific inferences such as [CAPTURED THE ROBBER] in SUPERMAN₁'s role, there will be several subjects (e.g. Verna) who do *not* share that concept. Therefore, we won't gain anything in terms of publicity by individuating concepts like SUPERMAN₁-SUPERMAN₂ locally rather than holistically!

Can Weiskopf's (PL) respond to (a)-(b) and save the localist proposal? I believe not. In response to (a), Weiskopf would claim that psychology should be our principled basis in selecting the individuating inferences. But I see no reason why our best psychological theory should include [CAPTURED THE ROBBER] in its description of SUPERMAN₁, while leaving [IS HEROIC] out. Moreover, suppose we do use psychology as our basis in individuating the concepts involved in contradictory beliefs: we will then be unable to account for those cases in which two concepts C₁-C₂ are described in *exactly the same way* by our best psychological theory (they are associated with the *same* prototype/set of exemplars/mini-theory), but are nevertheless involved in contradictory beliefs. Suppose for instance that, at t₁, I learn about an animal that is called “cat” in English and has certain prototypical features (meows, has whiskers, is four-legged etc...). Later, at t₂, I learn about an animal called “chat” in French that has those very same features; however, I also come to believe (perhaps because of reliable testimony) that the two terms designate different species which happen to have certain prototypical features in common. According to (FC), I must have two distinct concepts CAT and CHAT; on Weiskopf's view, however, the two concepts would not be distinct, since they

²⁸⁵ V. ch. 5 (sect. 4.2).

²⁸⁶ V. ch. 4 (sect. 4.3) and ch. 5 (sect. 4.2).

are associated with the same prototype²⁸⁷. This shows that (PL) is not fine-grained enough to individuate concepts as required by (FC) in every possible Frege case²⁸⁸.

(PL) is also vulnerable to (b). It's an immediate consequence of (PL)'s pluralism that several concepts are *not* shared. As Weiskopf notes (pp. 149-50), a doctor's cancer-concept has a complicated causal theory as its CC; therefore, it is not shared by Beth, whose cancer-concept has the CC [DISEASE MY UNCLE HAD]; nor is either of these concepts shared by Ada, who only knows of cancer as [DISEASE SMOKERS GET]. In addition, Weiskopf himself recognizes that many of the concepts involved in Frege cases are individuated very finely and not widely shared²⁸⁹. But if the localist agrees that such concepts are not shared, what motivation is left for not individuating them holistically?

Weiskopf might respond that this is not how psychologists individuate concepts: when providing a model for concept C, they only include in its content a *part* of the information that the subject associates with C (pp. 134-38). But there is no reason to use psychology as our principled basis in individuating *all* concepts! Locally structured representations might be especially useful for some theoretical purposes, such as explaining data about categorization, induction etc..., while being unsuitable for others, such as accounting for rationality in Frege cases. Weiskopf writes:

In general, when we have a model that is capable of accounting for some range of psychological phenomena, we should take it to be at least a provisional guide to the structure of the psychological mechanisms that underlie the production of those phenomena. [...] cognitive models that enjoy a significant degree of empirical validation are *prima facie* likely to be accurate descriptions of the psychological states, processes, and mechanisms underlying the task in question (p. 135).

I agree; but there might be psychological phenomena *other* than those investigated by psychologists, whose explanation requires assuming the existence of *different* mechanisms that are *not* correctly described by current empirical models. For instance, explaining rationality in Frege cases might require assuming that the subjects involved have coreferential concepts whose structure is not local, but holistic. These subjects will

²⁸⁷ The localist might reply that features like [CALLED “CAT”]-[CALLED “CHAT”] will be included in the two prototypes, thus making the concepts distinct. I find this reply *ad hoc*; at any rate, we could always devise a Paderewski-version of the case, where the same English term “cat” is employed at t_1-t_2 and I form the belief that two different species go by that same name.

²⁸⁸ V. Schiffer (1990). Interestingly, Weiskopf agrees:

[...] there are cases in which concepts appear to be distinct despite having identical cognitive and referential content (p. 154).

Weiskopf's solution is to claim that the concepts have different *formal properties*, since they are associated with distinct *mental files* (v. Recanati 1993, 2009). I cannot discuss the proposal here: note, however, that if files are individuated by their global inferential role Weiskopf's view will collapse into a form of holism; if not, we need to know what individuates files to establish whether the view can deal with Frege cases. (Weiskopf doesn't address the problem in his (2009a)).

²⁸⁹ P. 153.

then have *both* the local representations described by psychology *and* the holistic representations needed to satisfy (FC)²⁹⁰. Psychology is one guide to the structure of concepts, but it need not be the only one.

There is a second aspect of Weiskopf's view that I reject. Weiskopf takes the inferences which individuate a concept C to constitute a layer of C's *content*, which he calls "conceptual content" (CC). When conjoined with other aspects of his picture (v. points (1)-(3) *supra*), this gives rise to some bizarre consequences. Suppose it's part of my mini-theory of glass that glass is a solid. The CC of one of my glass-concepts will then include the feature [SOLID]. Now consider my thought IF SOMETHING IS A SAMPLE OF GLASS, THEN IT IS A SOLID²⁹¹. This thought expresses two different propositions, which are compositionally constructed from the CCs and the RCs of its components:

CC proposition: *If something is [C₁, C₂ ... C_n] and it is a solid, then it is a solid*

(where [C₁, C₂ ... C_n] is the set of features encoded in the CC of my concept GLASS). This proposition is *true*; indeed, it is *analytically* true on a "containment" conception of analyticity.

RC proposition: *If something is a sample of glass, then it is a solid*

This proposition is *false*, since glass is not a solid (it is a super-cooled liquid).

As Weiskopf recognizes, then, our thought is both *true* and *false*, since it expresses two propositions which are respectively true and false. In fact, the thought is both *analytically* true and (non-analytically) false, given the analytic nature of the CC proposition (pp. 156-60)!

Since a concept's CC does not fix its reference, many other thoughts will express analytically true CC propositions while also expressing false RC propositions²⁹². This problematic consequence would be avoided if we stopped taking CC to be part of a concept's *content*: there would then be no *proposition* expressed other than the RC one. Indeed, Weiskopf's decision to call CC "content" seems no more than an arbitrary terminological stipulation. If CC does not fix reference (and can sometimes be *entirely false* of the RC!²⁹³), what is it that makes it "content"? Surely, the fact that psychologists often talk this way doesn't give us sufficient reason to treat CC as genuine content, especially if this gives rise to the problems described in the previous

²⁹⁰ As usual, this claim will have to be declined differently depending on our ontology: a Fregean will postulate the existence of *different abstract concepts* (local and holistic), while a non-Fregean will simply claim that *different types* are involved. V. ch. 5 (sects. 4.1-4.2).

²⁹¹ Note that, as usual, I identify thoughts with mental representations (this follows Weiskopf's usage).

²⁹² This could not happen if the CC fixed the RC: v sect. 4 *infra*.

²⁹³ Pp. 142-43.

paragraph. My suggestion, then, is that, for those concepts that are individuated by a set of inferences (local or holistic), we simply take such inferences to be what individuates the concept and is required for its possession, without *also* taking them to constitute a further level of content over and above the concept's reference.

3. The Correctness Dilemma

3.1. A Dual Theory Of Concepts

In the last section, I examined a version of localism that responds to our first dilemma by simply identifying concepts with the structured mental representations posited by psychology. For most localists, however, this would be the wrong response to the dilemma: clearly, such localists would claim, not *all* of the mental representations that psychologists call “concepts” actually individuate corresponding concepts in the philosophical sense of the term²⁹⁴! The reason can be easily seen by looking at some well-known studies on categorization. For instance: in a famous experiment, Armstrong et al. (1981) showed that subjects find it natural to rank odd numbers for typicality, and that the typicality value assigned to a given number correlates with increased recognition speed and other analogous effects. This seems to show that subjects have a prototype representation for *odd number*. Surely, however, that representation does not have the same structure as the concept ODD NUMBER, which is presumably definitional!

Similar considerations apply to natural kind concepts: while we certainly have prototype or exemplar representations for skunks, these representations do not constitute the structure of the corresponding concepts. We do not judge painted raccoons to be skunks, even though they have strong superficial similarity with prototypical skunks²⁹⁵. Since we assign greater weight to “hidden” essential features than to superficial ones when categorizing these items, so the localist reasons, the structure of SKUNK should be identified with a mini-essentialist theory rather than a prototype or exemplar.

Clearly, once the localist rejects some of the mental representations posited by psychologists as non-individuating, our original “principled basis” problem will arise again: we now need a new criterion to select individuating inferences. The natural move to make at this point will be to identify individuating inferential dispositions with *fundamental* ones, thus adopting a “dual theory” of concepts²⁹⁶. On this view, the individuating inferences will be those that constitute the “core” of a concept. Conceptual cores are the “ultimate arbiters” of categorization: they include all and only those categorization rules that determine a subject’s categorization judgments under ideal conditions (unlimited time, complete information about the target object and so on). On the contrary, the “periphery” of a concept will be identified with those “rough and ready” categorization procedures that we employ when we have limited time or incomplete information. For instance, we often use prototypes/exemplars to make quick categorization decisions based on superficial properties, but our most reflective

²⁹⁴ V. e.g. Rey (1983, 1985) and Weatherson (2003), who both take this line.

²⁹⁵ Cf. Keil (1989).

²⁹⁶ Osherson and Smith (1981).

judgments will be based on whether an object has enough essential properties in common with other members of the category.

Most leading localist views can be seen as dual theories. For instance, on the localist view defended by authors like Rey and Peacocke, a subject S will have a concept C just in case he has that *implicit conception* which individuates C and is required for its possession²⁹⁷. An implicit conception is a subdoxastic state with propositional content. It consists of a set of (subpersonal) beliefs about the referent of the corresponding concept²⁹⁸: for instance, the implicit conception for the concept CHAIR will have the content [HAS A BACK, IS A SEAT...]; the implicit conception for DISJUNCTION will have the content [TRUE IFF AT LEAST ONE DISJUNCT IS TRUE]; and so on²⁹⁹. Possession of an implicit conception is analogous to possession of a competence with a certain syntactic rule, as in Chomsky's theory of "Universal Grammar"³⁰⁰. On Chomsky's view, we often make "performance errors" and fail to follow the subpersonal rules embedded in our language module. Still, we are "competent" with such rules, since we would follow them under idealized conditions where performance limitations were absent. Similarly, on the Rey-Peacocke view we are disposed to follow our implicit conceptions under ideal conditions, i.e. when we have enough time, information and so on. Under non-ideal conditions, however, we rely on rough and ready categorization procedures like prototypes or exemplars. Such procedures often determine categorization judgments that we would not make under ideal conditions: if I don't have any information about x 's essential properties, I might well judge it to be a skunk even though it is a painted raccoon, since it satisfies my skunk prototype. If I did have that information, however, I would follow my "essentialist" implicit conception and make the correct judgment.

As our brief overview makes clear, Peacocke and Rey's implicit conceptions are the "ultimate arbiters" of categorization: they are those fundamental rules that would determine our categorization judgments under ideal conditions³⁰¹. They can therefore be identified with the "conceptual cores" that individuate concepts and are required for their possession according to dual theories. My discussion of dual theories will thus take the implicit conceptions view as its main target, and I will talk interchangeably of cores and conceptions throughout: as the following sections will show, my arguments against the Peacocke/Rey view would also apply to other dual theories as such.

²⁹⁷ V. Peacocke (1998 a,b, 2007); Rey (2005, 2009b). Rey does not employ the notion of "implicit conception", but his picture can be assimilated to Peacocke's for the purpose of our discussion. V. also Jackson (1998), who seems to endorse both a dual theory and an implicit conceptions view.

²⁹⁸ Conceptions are sets of beliefs rather than inferences, but (as usual) we can move freely between talk of inferences and talk of beliefs. Given any inference of the form $X \text{ IS } F \rightarrow X \text{ IS } G$ which is allegedly required for possession of a concept C, that inference will correspond to the belief ALL FS ARE G: both the inference and the belief will thus be required for possession of the concept.

²⁹⁹ V. Peacocke (1998a), especially pp. 44-64.

³⁰⁰ V. Rey (2005), Peacocke 2007 (chs. 4-5).

³⁰¹ Rey (2009b) especially insists on this point.

3.2. Plausibility And Correctness: A Trade-Off

According to the localist view under examination, not *all* the structured mental representations posited by psychology count as concepts, but only the *fundamental* ones which constitute conceptual cores and form our subpersonal conceptions. Despite its popularity, this version of localism invites a second dilemma: if concepts are individuated by implicit conceptions, should we identify such conceptions with sets of *correct* inferences, or will *incorrect* inferences also be allowed in the individuating set?

We do have concepts: this, the localist cannot deny. A localist theory must not make the constraints on concept possession too stringent, or intentional agents will lack a number of concepts that they seem to employ all the time in their cognitive lives. So, if implicit conceptions are required for concept possession, they must be such that ordinary subjects can have them. They should not be identified with states that it is *psychologically implausible* to ascribe to ordinary thinkers, or we will make it too hard for them to have concepts at all.

Given this “psychological plausibility” constraint, we might now ask: *what are implicit conceptions?* Which of the subdoxastic states posited by our best psychological theories should be identified with them? For a large class of concepts, and particularly for *natural kind* ones³⁰², implicit conceptions will arguably have to be identified with “folk theories”³⁰³. Following standard doctrine in contemporary cognitive science, we can define folk theories as sets of basic theoretical principles that we employ with respect to a certain domain (physical objects, biological creatures, mental states and so on). Such theories are usually taken to be subpersonal and not immediately accessible to introspection. Moreover, they are standardly taken to be domain-specific and innate, thus bearing strong analogies to Chomsky’s Universal Grammar rules. They are also active in determining our most fundamental categorization procedures: for instance, the child’s disposition to consider hidden properties more important than superficial ones is presumably grounded in his innate folk biology. Finally, it’s psychologically plausible to ascribe such theories to ordinary thinkers, since there is strong independent evidence that they do play an important role in a number of cognitive processes³⁰⁴.

³⁰² In what follows I will mostly focus on natural kind concepts; it’s an interesting issue whether my arguments would also generalize to other kinds of concepts, but unfortunately I cannot address it here.

³⁰³ My overview of folk theories is largely based on Prinz (2002, ch. 8).

³⁰⁴ The psychologist’s use of the notion of a “folk theory” overlaps only in part with that of Jackson (1998). First, Jackson simply identifies folk theories with fundamental, subpersonal principles which drive our categorization judgments under suitably idealized conditions. Such principles need not be domain-specific and innate; in fact, they presumably won’t be for a large class of concepts such as artifact concepts. (Jackson might also question whether the folk theories for natural kind concepts like GOLD or GLASS would be domain-specific and innate, but this is too vast a topic to discuss here. In any case, none of my arguments in what follows turns on this point). Second, and more importantly, Jackson takes “folk theories” to fix the reference of the concepts they individuate, in which case they will be guaranteed to be

In sum, it seems that, if the constraints on concept possession imposed by a dual theory are ones that ordinary thinkers actually satisfy, implicit conceptions will have to be identified with folk theories for a large number of concepts. This raises our second dilemma for localism. The problem, in a nutshell, is that we have good reasons to think that our folk theories are often *wrong*. For instance, Spelke has argued that the “Aristotelian” principle “physical objects act on each other only on contact” is part of our folk physics, which is therefore inconsistent with Newtonian mechanics³⁰⁵. And our folk chemistry seems to include principles yielding the incorrect judgment that glass is a solid and not a liquid. So: *are mistaken folk-theoretical principles going to be part of the implicit conceptions which individuate our concepts?*

Theories of concepts that appeal to implicit conceptions, and dual theories more generally, will mostly answer this question in the negative: mistaken principles *cannot* be part of those fundamental implicit conceptions which are constitutively involved in our possession of the corresponding concepts³⁰⁶. Some passages from Peacocke do suggest a certain ambivalence on the topic:

It is not impossible for there to be an implicit conception with an incorrect content. A thinker may misunderstand some word in the public language. False presuppositions about certain kinds of objects or events in his environment may also enter the content of his implicit conceptions (Peacocke 1998a, p. 70).

Other passages, however, make clear that implicit conceptions cannot include false beliefs³⁰⁷:

An implicit conception is meant to be part of what it is, constitutively, to possess the concept involving that implicit conception. [...] if a principle involving a concept can intelligibly and correctly be abandoned, it cannot be any part of the content of an implicit conception involved in possessing that concept (Peacocke 1998b, p. 140).

It's easy to see why a localist will only allow true beliefs to be part of our implicit conceptions. The alternative view would be one on which:

- a) *All* concepts are individuated by the *fundamental* inferences which form our implicit conceptions/conceptual cores. Since non-fundamental inferences are not individuating, many of the structured mental representations described by the psychologist (e.g. prototypes/exemplars) do *not* count as concepts.

correct (v. sect. 4 *infra*); on the contrary, psychologists often identify folk theories with sets of false principles (v. next paragraph).

³⁰⁵ Spelke (1990).

³⁰⁶ I am not aware of anyone who explicitly takes implicit conceptions to include incorrect principles.

³⁰⁷ Peacocke confirmed this in conversation.

- b) At the same time, however, many individuating inferences are *incorrect*: for instance, GLASS → SOLID might be part of our implicit conception for GLASS.

The problem with this picture is that, once we allow for incorrect inferences to be individuating, the restriction to fundamental inferences becomes unmotivated. Why couldn't some concepts be individuated by *non-fundamental* sets of inferences, such as those included in a prototype? Clearly, the answer cannot be that, being employed under non-ideal conditions (limited time, information etc...), prototypes will often determine mistaken categorization judgments. While prototypes would indeed cause wrong judgments with respect to e.g. Keil's painted raccoons, our localist has already allowed for incorrect inferences to be part of conceptual cores (again, cf. GLASS → SOLID). Since these inferences will *also* determine mistaken judgments, the reason for counting prototypes as non-individuating can't be that they lead to bad categorization decisions.

Nor can the rationale be that, by only including fundamental inferences, we guarantee that concepts will be *shared* in all the cases covered by publicity. Consider for instance a chemist, who correctly identifies glass by its structural features rather than its superficial properties. Clearly, his concept of glass is not involved in the same fundamental inferences as ours: he *rejects* the inference GLASS → SOLID and accepts instead GLASS → LIQUID. Our localist must then concede that the chemist has a *different* concept GLASS*, one which (unlike ours) is associated with a correct conception³⁰⁸.

Finally, allowing for incorrect implicit conceptions would make dual theories unsuitable as a basis for "ambitious" conceptual analysis. Suppose we successfully analyze some concept C and discover the implicit conception associated with it³⁰⁹. Since that conception might well include incorrect principles, we still won't be able to draw any conclusions regarding the real-world category that is picked out by the concept. For instance, it would certainly be incorrect to conclude that, since GLASS → SOLID is part of our fundamental conception of glass, then glass must be a solid! If implicit conceptions can be incorrect, conceptual analysis will only yield knowledge about the structure of our concepts, not about the philosophical kinds that those concepts pick out. (The same is true of any theory on which incorrect inferences are individuating, such as Weiskopf's (PL). V. sect. 4 *infra* for more discussion).

These considerations show why localists like Peacocke generally hold that implicit conceptions must be *correct* if they individuate concepts. Choosing this horn of the

³⁰⁸ In fact, the restriction to fundamental inferences isn't simply unmotivated; there are also positive reasons to reject it. For one, I argued that those concepts which are involved in contradictory beliefs should be individuated holistically (ch. 5); but, clearly, many of the connections which individuate such concepts will *not* be fundamental. Moreover, I argued (chs. 2-3) that speakers will often "provide" certain structured concepts in their attitude ascriptions and intentional generalizations: but I see no reason to think that the inferences individuating such concepts will always be fundamental ones (are all the inferences included in the standard Clark-ish role [CALLED "CLARK", SHY, WEARS GLASSES...] fundamental?).

³⁰⁹ This picture of conceptual analysis is illustrated in detail by Peacocke (1998a, 2007).

dilemma, however, makes it hard to respect the psychological plausibility constraint. To fulfill that constraint, the localist will have to identify implicit conceptions with folk theories, at least for a large class of concepts. As we have seen, however, this will not be allowed if implicit conceptions are supposed to be correct, since folk theories are replete with errors!

Satisfying the correctness criterion would thus require selecting some set of correct individuating inferences *other* than folk-theoretical principles. But, of course, there is a trade-off between the correctness and the psychological plausibility constraints: it's hard to think what this set of correct inferential dispositions might be, such that it could plausibly be ascribed to ordinary thinkers. Suppose for instance that the conception individuating GLASS is that description of the kind *glass* that would be offered by a complete ideal chemical theory. Clearly, it would be psychologically implausible to hold that this conception was subpersonally cognized by ordinary thinkers! The localist must therefore hold that ordinary thinkers have GLASS in virtue of possessing some other implicit conception. The problem, however, is that it's hard to imagine what that conception could be, if we are not allowed to identify it with a (partially mistaken) folk theory.

A first possibility would be to appeal to deference³¹⁰. Even if concepts are individuated by correct implicit conceptions, someone might have a concept while having an incorrect conception of its referent. This is because, despite his imperfect conception, the subject might *defer* to the experts in his linguistic community: I can have ARTHRITIS by deferring to the experts, even if my conception of arthritis is partially wrong³¹¹.

The appeal to deference appears problematic for several reasons, but I will focus on one. At least when it comes to natural kinds³¹², it seems uncontroversial that all the existing experts in a community could be (and often are) wrong about the properties of a certain specific kind. In this scenario, how could the experts and the folk possess the relevant concepts? If these two classes of thinkers do have the concepts in question, then *ex hypothesi* they both do so in virtue of having mistaken conceptions about the corresponding kinds. But then mistaken conceptions can be part of the possession conditions for concepts and we are back to the first horn of the dilemma.

A more convincing response would be that not *all* components of folk theories are incorrect. Within each folk theory, there is a “core” subset of correct inferences surrounded by a periphery of mistaken principles: this core is what individuates our concepts of the corresponding kinds. And since cognitive science offers plenty of independent evidence in favor of folk theories, the psychological plausibility constraint is met: these are conceptions that ordinary thinkers can plausibly be taken to have.

But why should we think that folk theories will embed correct conceptions about the corresponding kinds? After all, we have seen that many folk principles have been

³¹⁰ Cf. Peacocke (1992, ch. 1).

³¹¹ Burge (1979).

³¹² But v. Burge (1986) for a case involving artifact kinds.

empirically falsified, so the localist must provide some reason to think that this will not happen with *all* the beliefs that constitute our theories. In response, our localist could take the “core” conception within a folk theory to *fix the reference* of the corresponding concepts. Since core conceptions determine the conditions for something to fall under concept C, they will by definition be true of C’s extension. Of course, some of the principles that the psychologist includes in our folk theories will not be part of the reference-fixing core (the set of principles according to which glass is a solid presumably won’t be), but this is compatible with there being such a core.

In conclusion: if folk theories include reference-fixing inferences, we can *both* make a localist theory of concept possession psychologically plausible (there is independent evidence that ordinary thinkers have folk theories) *and* take concepts to be individuated by correct inferences (reference-fixing inferences are guaranteed to be correct)³¹³. This move would thus enable the localist to answer our dilemma; whether it can be made to work will be my topic in the next section.

³¹³ Indeed, on the first version of Peacocke’s theory, individuating inferences were explicitly taken to be reference-fixing and were thus guaranteed to be truth-preserving (v. Peacocke 1992, pp. 16-24).

4. Reference-fixing conceptions?

In this section, I will focus on those versions of localism on which *all* the inferences which individuate concepts and are required for their possession are *correct* (truth-preserving). This is undoubtedly the majority view among localists, and I will refer to it as “true localism”.

There are several reasons behind the popularity of true localism. Notice, in particular, that *only* a true localist view will warrant an ambitious version of the conceptual analysis project. If the conception that individuates concept C is guaranteed to be correct, we will be able to draw substantive conclusions about C’s reference once we find out what that conception is. Having discovered through a process of conceptual analysis that BACHELOR is individuated by [UNMARRIED MAN], the true localist can validly conclude that something is a bachelor just in case it is an unmarried man. If individuating conceptions were often incorrect, that conclusion would be unwarranted.

Of course, the attractive features of true localism come at a cost. As we have seen, the view raises a problem of psychological plausibility: *what* are these correct inferential dispositions that can be plausibly ascribed to ordinary thinkers and allow them to possess the corresponding concepts? We noted in the last section that, for a number of concepts, the localist will have to identify such dispositions with *reference-fixing conceptions* embedded in our folk theories. The problem, of course, is spelling out what these conceptions are in a psychologically plausible way: reference-fixing conceptions must be such that we can reasonably take ordinary thinkers to have them.

In this section, I will consider a cluster of related objections that any true localist view will have to face. I will argue that true localism cannot answer such objections, and more specifically that it won’t be able to do so while providing a psychologically plausible account of reference-fixing conceptions. My examination of localism will then be completed in the next section, where I will make some more general remarks about the various localist pictures we have examined and about my own theory of concept individuation/possession.

My arguments in this section will target one of the best-developed true localist theories on the market, the “descriptivist” picture of concepts and mental content recently developed by Frank Jackson³¹⁴. According to Jackson, concepts are individuated by *reference-fixing descriptions* which determine their reference. For natural kind concepts, these take the form of rigidified descriptions: WATER, for instance, might be individuated by the description [THE ACTUAL WATERY STUFF OF OUR ACQUAINTANCE]. Jackson’s reference-fixing descriptions are conceptions required for possession of the corresponding concepts: a thinker doesn’t have WATER unless he accepts $X \text{ IS WATER} \leftrightarrow X \text{ IS THE ACTUAL WATERY STUFF OF OUR ACQUAINTANCE}$.

³¹⁴ V. especially Jackson (1998). While I will focus on Jackson’s view, my arguments would also apply to other true localist theories.

Jackson's view is thus an excellent example of true localism: those bodies of beliefs that the psychologist calls "folk theories" have a core, correct part which fixes the reference of the corresponding natural kind concepts and is required for their possession.

In what follows, I will argue that true localist pictures such as Jackson's will make implausible predictions about the reference of natural kind concepts in all those cases where someone does *not* accept the description which allegedly fixes the reference of some concept *C*, and yet seems able to refer to the natural kind picked out by *C*. An objection raised by Schroeter (2004) against Jackson will serve as our starting point. Having described Jackson's view, Schroeter asks: what exactly will a reference-fixing description for natural kind concepts look like? Arguably, the description should consist of two elements: an *actual world description* and a *sortal*. To see why, consider again the toy-description [THE ACTUAL WATERY STUFF OF OUR ACQUAINTANCE], which Jackson identifies with our concept WATER. This description is not sufficiently specific, as it doesn't specify what sortal is being employed to fix the reference. What *kind* of watery stuff are we referring to? If it was just the actual kind comprising all and only the liquids with watery features, Putnam's XYZ would also fall under the concept: we *are* actually acquainted with the kind including all and only the watery substances, since we are actually acquainted with local samples of H₂O. So what we need to get the right extension is something like: [WHATEVER NATURAL KIND/CHEMICAL KIND IS THE ACTUAL WATERY STUFF OF OUR ACQUAINTANCE].

Unfortunately, the need for a specific sortal raises a problem: for many subjects, the only sortal that they can plausibly be taken to employ is not satisfied by anything in the actual world:

Consider the four classical elements: earth, air, fire and water. Aristotle took these to be four basic configurations of prime matter which entered into the constitution of all material bodies. [...] So [Jackson] might be tempted to say Aristotle's "water" concept referred to that basic configuration of prime matter which most closely satisfied Aristotle's own criteria for identifying water in the actual world. If that is what Aristotle had in mind, however, his "water" concept did not manage to refer to anything at all – or at least not anything in this world (Schroeter 2004, p. 437).

The reference-fixing description that Aristotle seems to associate with his concept (call it WATER*)³¹⁵ is something like: [THE BASIC CONFIGURATION OF PRIME MATTER WHICH IS THE ACTUAL WATERY STUFF OF OUR ACQUAINTANCE]. Clearly, that description does not pick out anything in the actual world; it picks out a non-instantiated kind *water** which

³¹⁵ Notice that Aristotle's concept is *not* associated with our description [WHATEVER NATURAL KIND/CHEMICAL KIND IS THE ACTUAL WATERY STUFF OF OUR ACQUAINTANCE], but with a different description. Jackson will therefore take Aristotle's WATER* to be a *different concept* from our WATER. I think this would be an acceptable consequence of the view, as long as Aristotle's concept and ours had *same reference* (v. sect. 5 *infra* for discussion of this point).

is *not* the chemical kind *water* (= H₂O) to which our concept WATER refers. This creates several problems³¹⁶:

1) *Aristotle was not thinking and talking about water, i.e. about H₂O.*

Intuitively, it seems that Aristotle was thinking and talking about the same kind *water* as us (Schroeter 2004, p. 437). But if the reference of his WATER* was fixed by the description above, that concept had a different reference from our WATER. Moreover, since Aristotle didn't have any other concepts whose reference could be fixed to *water* via a description close or identical to ours, he simply had no way to express propositions about water in thought or language; he could only think/speak about the non-existent *water**.

2) *Aristotle was not disagreeing with us about water, and he was not mistaken about its nature.*

The first problem gives rise to further issues having to do with disagreement and error. Aristotle's theory of water, according to which it is the basic configuration of prime matter which plays the watery role in the actual world, seems to be inconsistent with our theory of water; moreover, it would appear to be a wrong theory³¹⁷. But if the description above was fixing the reference of Aristotle's WATER*, his theory was *true*, only not about *water* but about the non-existing *water**. Nor was Aristotle *disagreeing* with us about the nature of water, i.e. of H₂O: he was simply talking about some other, non-existing kind *water**.

The localist might reply that *there is* a proposition which is respectively accepted and rejected by Aristotle and me, namely the one expressed by the following existential statement:

(E) $\exists x$ (x is a sample of the basic configuration of prime matter which is the actual watery stuff of our acquaintance)

When Aristotle asserts (the Greek equivalent of) (E), he expresses a proposition which is *inconsistent* with the one expressed by my utterance of:

(En) $\neg \exists x$ (x is a sample of the basic configuration of prime matter which is the actual watery stuff of our acquaintance)

³¹⁶ Objection (1) is raised by Schroeter herself. Schroeter is also worried about the problem of error in (2), but she does not talk about disagreement and does not consider the possible descriptivist reply discussed below. She also doesn't raise (3).

³¹⁷ Schroeter (2004, p. 437).

Since a common propositional content is being asserted and denied, there is genuine disagreement. What we do *not* disagree about is the proposition expressed by the universally quantified statement:

(U) $\forall x (x \text{ is a sample of water just in case it is a sample of the basic configuration of prime matter which is the actual watery stuff of our acquaintance})$

Aristotle's utterance of (U) is true: his “water” is synonymous with his reference-fixing description, so his utterance expresses the following proposition (P):

(P) $\forall x (x \text{ is a sample of } \text{the basic configuration of prime matter which is the actual watery stuff of our acquaintance} \text{ just in case it is a sample of the basic configuration of prime matter which is the actual watery stuff of our acquaintance})$

On the other hand, my utterance of (Un) (= the negation of (U)) is *also* true, since my “water” is synonymous with a *different* description, one which picks out H₂O³¹⁸.

A similar move is available for error. Aristotle was wrong when making the existential claim (E), which does express a false proposition in his mouth as well as ours; on the other hand, he was *not* wrong when asserting (U), which is indeed true of the kind *water** that his term “water” allegedly picks out.

This reply is not fully satisfactory. The localist position clashes with our intuition that Aristotle had a theory *of water* (i.e. of the same kind we are referring to!), and that his theory was refuted by later empirical discoveries showing that there is no such thing as prime matter. Those discoveries didn't just show that Aristotle's assertion of (E) was false, they also showed that his assertion of (U) was³¹⁹; similarly, there was genuine disagreement between the modern chemists who made those discoveries, asserting (Un) as a result, and Aristotle asserting (U).

To make the worry more vivid, suppose Aristotle learns about recent empirical findings regarding the chemical composition of the watery stuff on Earth and the non-existence of prime matter. Suppose he reacts as follows (a likely reaction, in fact): “I

³¹⁸ I am ignoring issues having to do with translation from Ancient Greek into English. I also assume that my arguments would still go through if (E)/(U) were mental representations rather than natural language sentences; for instance, (E)/(U) might be sentences in Aristotle's LoT. (Jackson (1998) seems to accept the assumption, which is explicitly endorsed by other “two-dimensionalists” such as Chalmers (2002, 2006)).

I will keep moving between the level of thought and the level of language in this section. Of course, one might hold that there are significant differences between linguistic and mental content: in particular, one could take linguistic content to be public and shared, due to its conventional nature, while taking mental content to be essentially private and often not shared. I don't think this would help Jackson in Aristotle's case, but this is clearly too vast a topic to be discussed here.

³¹⁹ Notice that Aristotle's assertion of (U) is *not* existentially committing as to the existence of something which satisfies the description he associates with “water”. Therefore, the localist cannot account for error by claiming that (U) is false if (E) is.

guess I was wrong when I asserted (U); water is not a basic configuration of prime matter, but a chemical kind. Water does exist, of course; it's just not what I thought it was". It seems hard for the descriptivist to account for Aristotle's reaction: assuming Jackson's theory, Aristotle should simply conclude that *there is no water!* Moreover, we will now have to take Aristotle to be wrong about the content of his own assertions: if descriptivism is true, then Aristotle asserted a *true* proposition with (U), despite his current protests to the contrary!

3) *Aristotle's WATER-thoughts and "water"-sentences are either false or lack truth-value*

Suppose Aristotle asserts: "Water is drinkable" (or thinks the corresponding thought). Since "water" is synonymous with the definite description "the basic configuration of prime matter which is the actual watery stuff of our acquaintance", Aristotle's utterance will be true just in case the basic configuration of prime matter which is the actual watery stuff of our acquaintance is drinkable. But then, depending on your preferred account of definite descriptions, his utterance will be either false or neither true nor false, since no existing kind satisfies this description. On a *Russellian* analysis of definite descriptions, the utterance is *false*, since it entails the existential claim:

(EC) $\exists x$ (x is the basic configuration of prime matter which is the actual watery stuff of our acquaintance)

On a *Fregean* account, Aristotle's utterance is *neither true nor false*, since (EC) is only presupposed by it. On both accounts, Jackson's view yields the wrong prediction, since Aristotle's utterance seems plainly true: moreover, the problem will extend to countless other claims/thoughts ("water is transparent", "there is water in the glass"....).

To answer (1)-(3), the localist will probably try to identify a more "fundamental" description which was shared by us and Aristotle and which fixed the reference of his WATER* to H₂O. But what could that description be? Suppose we drop the problematic sortal [PRIME MATTER] and go instead for: [STRUCTURAL KIND WHICH IS THE ACTUAL WATERY STUFF OF OUR ACQUAINTANCE]. Now:

Aristotle also thought earth could be explained in terms of its fundamental structural features. We think he was wrong. The category of earth, we think, ought to be explained in terms of the practical concerns of farmers and engineers. There is no unified structural kind underlying different samples of earth (Schroeter 2004, p. 238).

Aristotle thought the four elements were exactly on a par: he took *each* of them to be individuated by its "deep, structural" features (i.e. by its being a certain configuration of prime matter). So, if the reference of Aristotle's WATER* was fixed by a description including the sortal [STRUCTURAL KIND], the reference of his EARTH* was presumably

fixed by a description including the same sortal. But then the descriptivist will have to face objections (1)-(3) again, only with respect to EARTH* rather than WATER*. For we have found out empirically that there are no unifying structural features underlying the samples of what Aristotle would have called “earth”: *earth* is a superficial, functional kind individuated by its relations to human concerns. Therefore, Aristotle’s concept EARTH* will now fail to pick out any existing kind, it will have a different reference from our EARTH, and objections (1)-(3) will arise again.

Maybe our conceptions of water and earth still have a sortal in common with Aristotle’s, since he thought of both of them as *substances*? But, Schroeter notes, Aristotle also thought *fire* was a substance, while we have found out empirically that it is a process. For any sortal which was plausibly part of Aristotle’s reference-fixing descriptions, we can think of some empirical discovery which showed that the sortal in question was *not* satisfied by the kind that, nonetheless, Aristotle was intuitively talking about.

Schroeter stops here in her ingenious list of counterexamples, but the localist is not done yet. At this point, he might try to “shrink” Aristotle’s reference-fixing description for water to: [WHATEVER THOSE THINGS (pointing at watery samples) ARE]. The indexical-causal link would then connect Aristotle to the chemical kind *water/H₂O* underlying the local watery samples, despite his mistaken metaphysical views about what underlies those samples. The problem with this suggestion is that it massively underdetermines reference. In his demonstration of watery samples, Aristotle was in causal-indexical contact with many different kinds: *H₂O*, *liquid*, *transparent*, *watery*... Which of these constitutes the reference of his concept? To save the proposal, we would need to add a more specific sortal to the demonstrative in the description, but this would of course simply raise the original problem again.

At this point, the localist might decide to employ a completely different reply. Let the “*anti-descriptivist*” (AD) be someone who denies that a concept’s reference is fixed by a description which is required for possession of that concept. Now, the descriptivist might observe, AD must propose *some* metasemantic theory alternative to descriptivism; roughly speaking, he must tell some story about how our concepts come to have the reference they intuitively have, a story that doesn’t appeal to reference-fixing descriptions at any point. Of course, there are several AD theories that could be exploited for this purpose. According to *teleological* theories, for instance, my concept WATER picks out H₂O (among all the eligible candidates) because it is its “proper function” to refer to H₂O³²⁰; according to Fodor’s *asymmetric-dependency* theory, my concept WATER stands in a relation R of reliable covariation with H₂O, and even though the concept also stands in analogous relations with other natural kinds, R is “special” since such relations would not obtain if R didn’t also obtain³²¹.

³²⁰ Cf. Millikan (1984, 1993).

³²¹ Cf. Fodor (1987, 1990).

Other AD theories have been defended, but what matters for our purposes is this: *for any anti-descriptivist metasemantics, the localist can propose a corresponding description which incorporates that metasemantics*. Let “T” be an AD theory about what fixes the reference of Aristotle’s WATER*. The descriptivist will then take Aristotle’s reference-fixing description for WATER* to be: [THE WATERY STUFF OF OUR ACQUAINTANCE THAT SATISFIES T]. For instance: if T is teleological, the description will be: [THE WATERY STUFF OF OUR ACQUAINTANCE THAT IT’S THE PROPER FUNCTION OF MY CONCEPT ‘WATER*’ TO REFER TO]³²² (*mutatis mutandis* for other anti-descriptivist views).

This move seems to show that the anti-descriptivist *cannot* successfully raise objections (1)-(3) against Jackson. Clearly, an anti-descriptivist who raises (1)-(3) will take them to be sound; therefore, he will also agree that, if his own metasemantic theory T is correct, it must entail that Aristotle’s WATER* refers to H₂O (or (1)-(3) will arise). But if T entails that WATER* refers to H₂O, a description *including* T will *also* pick out H₂O! All the descriptivist needs, then, is to claim that, for any theory T proposed by his opponent, Aristotle’s reference-fixing description for WATER* will include T: WATER* will then refer to H₂O, and (1)-(3) will not arise³²³.

The move is subtle, but there are good reasons to resist it. While it is indeed *possible* to incorporate any metasemantic theory in a reference-fixing description, it is a different matter to show that the description *was actually employed* in fixing the reference. In particular, the move seems to blatantly violate the psychological plausibility constraint. The localist needs to postulate that ordinary thinkers like Aristotle routinely employ descriptions which incorporate *precisely* that meta-semantic theory T which avoids (1)-(3). But, first, the move appears entirely *ad hoc*; we have no independent psychological evidence showing that people take the reference of their concepts to be fixed by e.g. a teleological or an asymmetric-dependence mechanism. Second, the metasemantic theory needed to avoid the objections might be so complicated that it couldn’t plausibly be ascribed to ordinary thinkers.

The localist might reply that thinkers do have intuitions about possible cases showing what descriptions they associate with their concepts. For instance, our intuitions about Kripke’s Gödel–Schmidt case seem to show that the description we associate with the name “Gödel” (and with the corresponding singular concept) must

³²² Imagine this to be something like a metalinguistic description used by Aristotle to introduce the term “water” for the first time.

³²³ Jackson himself suggests a similar move (v. Jackson 2007a, especially pp. 19-20). For instance, here is how he incorporates causal theories of reference in his descriptivist semantics for proper names:

[...] we should expect the most plausible candidate for the associated description [...] for a name “A” to be, in many cases, something like *the thing of such and such a kind at the far end of an information-preserving causal chain ending in a certain token of “A” in a certain sentence token of, say, the form “A is F”* (Jackson 2007b, p. 249).

include a “causal” element³²⁴. So we do have independent evidence supporting the claim that Aristotle’s description for WATER* includes theory T (however we decide to fill “T” in). And, since our associating a certain description with a concept might be a subpersonal state analogous to our “cognizing” a complicated syntactic rule, T can be as sophisticated as needed³²⁵.

In reply, notice that, if the localist identifies T with one of the AD metasemantic theories that have historically been proposed, it’s very unclear that *all* ordinary thinkers will have intuitions conforming to it. For instance, teleological theories of content notoriously yield counterintuitive results in certain cases, and defenders of such views have often decided to simply reject such intuitions as incorrect³²⁶. If on the other hand the localist claims that existing theories are not satisfactory, and that the metasemantic theory T that will be incorporated in the reference-fixing description is yet to come, then we just don’t know yet whether ordinary thinkers will have intuitions confirming that they implicitly believe the theory. *Prima facie*, it’s plausible to hypothesize that at least some people’s intuitions will not accord with this future ideal metasemantics. If this turns out to be the case, then the localist will either have to maintain that such subjects *would* have the intuitions under certain suitably idealized conditions, *or* that the reference of their concepts is *not* fixed by a description which incorporates T. If the localist chooses the first option, some empirical evidence must be provided to show that the subjects would indeed have the intuitions under ideal conditions (where the conditions in question should not make the theory non-falsifiable: if we idealize enough, everyone will have the “right” intuitions...). If he chooses the second, he’ll concede that these subjects don’t share reference with us, in which case objections (1)-(3) will arise again.

This brings us to a second problem with the descriptivist’s move. To see what’s wrong with that move, we don’t even need to show that *actual* thinkers might fail to accept theory T; appealing to *possible* thinkers will be enough. Suppose someone (“Aristotle*”) does *not* seem to take the reference of one of his natural kind concepts to be fixed by the description: [THE WATERY STUFF OF OUR ACQUAINTANCE THAT SATISFIES T] (where, recall, T is any AD metasemantic theory which would entail that WATER* refers to H₂O). In the imagined case, Aristotle* doesn’t seem to employ the description at *any* level, be it personal or subpersonal (e.g. if T is teleological, Aristotle* will have intuitions about possible cases that are inconsistent with teleological theories; he will stick to them under ideal conditions; he will explicitly hold a non-teleological theory of content, etc...). The case is otherwise identical to our original scenario involving Aristotle: in particular, Aristotle* also identifies the referent of his “water” with a configuration of prime matter underlying the watery stuff of our acquaintance. Here, none of the options available to the localist seem at all plausible:

³²⁴ V. Jackson (1998, chs. 2-3); Jackson (2007b, p. 249).

³²⁵ Jackson (1998, pp. 34-38).

³²⁶ Cf. Millikan (1993, p. 149) on Pietroski’s (1992) “kimu-snorf” case.

- The localist could *deny* that his prediction about the case is incorrect; we *should* treat someone like Aristotle* as referring to a non-existing kind *water**. But, if objections (1)-(3) apply in the case of Aristotle, they should *also* apply in the case of Aristotle*, since the two scenarios appear exactly analogous.
- Alternatively, the localist might try to find a reference-fixing description that can be plausibly ascribed to Aristotle* and picks out H₂O. But all the candidates we examined for Aristotle failed, and the appeal to T is ruled out by hypothesis.
- Finally, the localist might argue that the case is for some reason impossible, but I find it hard to imagine how such an argument might go.

I conclude that including theory T in the reference-fixing description for Aristotle's WATER* will not constitute a satisfactory reply to objections (1)-(3).

5. Conclusion

In this chapter I have examined different varieties of localism about concept individuation and possession. First, I have discussed pluralist-localist views that identify concepts with the locally structured representations posited by cognitive psychology (sect. 2.3). Such views have some interesting features in common with my own pluralist picture; however, they are insufficiently radical and should extend their pluralism to allow for holistically individuated concepts, as well as other concepts which will not be individuated in the way prescribed by the psychologist.

I have then considered localist views on which not all the structured representations described by psychology will count as “concepts”, but only the fundamental ones that drive our categorization decisions under certain ideal circumstances (sect. 3). Such views must deal with the further dilemma of whether to allow for fundamental, but incorrect dispositions to individuate concepts. I considered a few reasons why localists seem to uniformly reject this possibility, and then noted that this creates a problem of psychological plausibility (sect. 3.2): if only correct inferences are individuating, we must find (for each concept) a corresponding correct conception that can plausibly be ascribed to ordinary thinkers. This is not easy, since our “folk theories” will often include mistaken principles.

Presumably, the localist’s response will be that (parts of) our folk theories *fix the reference* of our concepts, so that they are guaranteed to be correct. I have then examined a “true localist” picture on which concepts are individuated by such reference-fixing conceptions (sect. 4), arguing that it will end up making mistaken predictions about subjects like Aristotle. These subjects have “deviant” implicit conceptions: therefore, if their deviant conceptions fix the reference of their concepts, such concepts will turn out to have different reference from ours. As we have seen, however, this is an unacceptable consequence, since it gives rise to objections (1)-(3). I have then considered and rejected some possible localist replies. In particular, I focused on Jackson’s suggestion that we can always include an anti-descriptivist metasemantics in our reference-fixing descriptions; in response, I argued that the move can’t deal with subjects like Aristotle*, and that in any case it will again expose the localist to a problem of psychological plausibility.

I would like to conclude with a few additional remarks about “deviant subjects”, an issue which was raised at the very beginning of our discussion³²⁷. What’s the *real* problem with anomalous thinkers like Aristotle? Many defenders of concept publicity believe the problem is that, if a localist IRS theory was true, deviant subjects would not have the *same concepts* as us. Here is a revealing passage from Laurence and Margolis:

³²⁷ V. my summary of Williamson (2007) in the general Introduction.

Consider the situation of someone who grows up with no exposure to lakes or oceans – maybe a nomad in the Sahara. If Jackson’s analysis is right, such a person should be able to know a priori that water fills oceans and lakes *or else she lacks the concept WATER altogether*³²⁸. We take it, however, that neither of these possibilities is at all plausible [...] (Laurence and Margolis 2003, p. 262, my emphasis).

If localism is true, subjects like the nomad or Aristotle won’t have the same concept WATER as us, because of their anomalous inferential dispositions. Notice, however, that this objection might not be particularly effective against certain localist views. An “untrue” localist like Weiskopf could respond that, while deviant subjects do not share our concept WATER, their concept WATER* still has same *reference* as our WATER. This is because, according to untrue localists, the inferences which individuate a concept *do not* fix its reference³²⁹. Therefore, the nomad or Aristotle can still have a concept WATER* that refers to the same chemical kind as our WATER, even though their inferential dispositions are radically different from ours. Our localist will then claim that publicity is preserved, since sameness of reference is all that’s needed to account for phenomena like communication, agreement/disagreement and so on³³⁰.

In light of my arguments in previous chapters, I take a (much qualified) version of this response to be substantially correct. We *do* share *some* of our water-concepts with the nomad and Aristotle: these are those water-concepts that are individuated by their reference, or by a small set of inferential connections. At the same time, however, deviant subjects do not share *all* of our water-concepts: for instance, they clearly don’t have a concept WATER that is associated with the local inferential role [ACTUAL WATERY STUFF OF OUR ACQUAINTANCE]. (There will also be further concepts which are not shared, e.g. holistically individuated water-concepts). Luckily, on the picture of concepts I developed in chapter 5, sharing of the more coarse-grained concepts is all that’s required to account for publicity.

Deviant subjects become much more worrying once we take reference to be fixed by our inferential dispositions, as on Jackson’s “true localist” picture. Because their inferential dispositions are so different from ours, these subjects will now refer to *different natural kinds* from us³³¹. For instance, Aristotle will not only have a different concept WATER*: more worryingly, that concept will also have different *reference* from our WATER, since the associated inferences will pick out the non-instantiated kind *water**. This, in turn, will give rise to objections (1)-(3). Notice, in particular, that there is now *no level* where the theory can account for phenomena like

³²⁸ (In Jackson’s reference-fixing description for WATER, [WATERY] is glossed as including [FILLS OCEANS AND LAKES] as well as the other standard watery features).

³²⁹ Recall, for instance, that Weiskopf takes reference to be fixed by “externalist” mechanisms like causal chains or deference (sect. 2.3). Block (1993, p. 56) makes the same move.

³³⁰ This is the line of Prinz and Clark (2004) and Block (1993).

³³¹ This problem might not arise with Laurence and Margolis’ nomad: his concept could still pick out water/H₂O, but through a description that is partially different from ours.

agreement/disagreement or communication with Aristotle: concepts are not shared, but reference is not shared either.

Deviant subjects might well fail to have some of our concepts, as long as we can still think and talk about the same objects and properties. If the inferences which individuate a concept also fix its reference, however, neither of these *desiderata* will be met. For this reason, it's an important part of my picture that reference must *not* be inferentially fixed³³². Of course, this creates a dialectical obligation: I have to provide a non-inferentialist metasemantics alternative to Jackson's and show that this metasemantics will make correct predictions about subjects like Aristotle. This, however, must be a topic for some other time.

³³² Because conceptual structure is not reference-fixing, I'm also skeptical about "ambitious" versions of the conceptual analysis project: an analysis of the structure of our concepts will *not* yield any knowledge about the nature of the philosophical kinds picked out by those concepts (v. *supra* for more details about ambitious/non ambitious conceptual analysis).

Appendix A

Chapter 1: List of numbered sentences

- 1) Superman can fly.
- 2) Clark Kent can fly.
- 2n) Clark Kent cannot fly.

- 3) Lois Lane believes that Superman can fly.
- 4) Lois Lane believes that Clark Kent can fly.
- 4n) Lois Lane doesn't believe that Clark Kent can fly.

- 5) Lois Lane believes the proposition that Superman/Clark Kent can fly under the guise of the sentence "Superman can fly".
- 6) Lois Lane believes the proposition that Superman/Clark Kent can fly under the guise of the sentence "Clark Kent can fly".

- 7) Lois Lane believes that Superman, the mighty superhero, can fly.
- 8) Lois Lane believes that Clark Kent, the milquetoast reporter, can fly.

- 9) Superman is Clark Kent.

- 10) The sentences "Lois Lane believes that Superman can fly" and "Lois Lane believes that Clark Kent can fly" express different propositions.

- 11) If Lois Lane is rational, reflective, and attentive, and she believes that Clark Kent cannot fly, then she doesn't also believe that Clark Kent can fly.

- 12) If a person wants Twain to autograph her book, and she believes that if she waves then Twain will autograph her book, then, other things being equal, she will wave.

- 13) If a person wants Twain to autograph her book, believes that if she waves then Clemens will autograph her book, and doesn't believe that Twain is Clemens, then, other things being equal, she will not wave.

- 14) If a person wants Twain to autograph her book, believes that if she waves then Clemens will autograph her book, and wants and believes these propositions in mismatching ways, then, other things being equal, she will not wave.

Appendix B

Chapters 2-3: List of numbered sentences

G) If a subject S wants P and believes that if she performs action a then P, then other things being equal S will perform action a .

G1) If a subject S wants to get water, then other things being equal S will look for water.

G2) If a subject S wants to get water and believes that if she opens the fridge she will get water, then other things being equal S will open the fridge.

G3) If a subject S wants to get orange juice and believes that if she opens the fridge she will get orange juice, then other things being equal S will open the fridge.

G4) If a subject S wants to annoy x and believes that if she opens the fridge she will annoy x , then other things being equal S will open the fridge.

G5) If a subject S believes that Clark can fly and believes that Clark just jumped off a skyscraper, then other things being equal S will not be worried.

G6) If a subject S believes that Superman is nearby and believes that Lex Luthor is nearby, then other things being equal S will run towards Superman.

G7) If a subject S does not believe that Superman is nearby and believes that Lex Luthor is nearby, then other things being equal S will run away.

(L) Lois believes that Clark can fly.

(L_s) S believes that Clark can fly.

References

- Arjo, D. (1996), "Sticking Up for Oedipus: Fodor on Intentional Generalizations and Broad Content", *Mind and Language* 11 (3), pp. 231-245.
- Armstrong, S. L. et al. (1983), "What Some Concepts Might Not Be", *Cognition* 13, pp. 263-308.
- Aydede, M. (1998), "Fodor on Concepts and Frege Puzzles", *Pacific Philosophical Quarterly* 79 (4), pp. 289-294.
- Aydede, M. (2000a), "On the Type/Token Relation of Mental Representations", *Facta Philosophica: International Journal for Contemporary Philosophy* 2 (1), pp. 23-49.
- Aydede, M. (2000b), "Computation and Intentional Psychology", *Dialogue* 39 (4), pp. 365-379.
- Aydede, M. and Robbins, P. (2001), "Are Frege Cases Exceptions to Intentional Generalizations?", *Canadian Journal of Philosophy* 31 (1), pp. 1-22.
- Barsalou, L. W. (1989), "Intraconcept Similarity and its Implications for Interconcept Similarity", in S. Vosniadou & A. Ortony (eds.), *Similarity and Analogical Reasoning*, Cambridge: Cambridge University Press, pp. 76-121.
- Barsalou, L.W. (1993), "Challenging assumptions about concepts", *Cognitive Development* 8, pp. 169- 180.
- Bealer, G. (1998), "Intuition and the Autonomy of Philosophy", in M. DePaul & W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman & Littlefield, pp. 201-239.
- Block, N. (1993), "Holism, Hyper-Analyticity and Hyper-Compositionality", *Philosophical Issues* 3, pp. 37-72.
- Block, N. (1995), "An Argument for Holism", *Proceedings of the Aristotelian Society* Vol. XCIV, pp. 151-169.
- Block, N. (1998), "Holism, Mental and Semantics", in *The Routledge Encyclopedia of Philosophy*.

Boghossian, P. A. (2001), "How Are Objective Epistemic Reasons Possible?" *Philosophical Studies* 106 (1-2), pp. 1 - 40.

Boghossian, P. A. (2003), "Epistemic Analyticity: A Defense", *Grazer Philosophische Studien* 66 (1), pp. 15-35.

Braun, D. (1998), "Understanding Belief Reports", *Philosophical Review* 10 (7), pp. 555-595.

Braun, D. (2000), "Russellianism and Psychological Generalizations", *Noûs* 34, pp. 203-236.

Braun, D. (2001a), "Russellianism and Prediction", *Philosophical Studies* 105, pp. 59-105.

Braun, D. (2001b), "Russellianism and Explanation", *Philosophical perspectives* 15, pp. 253-289.

Braun, D. (2006), "Illogical, but Rational", *Noûs* 40, pp. 376–379.

Braun, D. and Saul, J. (2002), "Simple Sentences, Substitution, and Mistaken Evaluations", *Philosophical Studies* 11 (1), pp. 1-41.

Braun, D. and Sider, T. (2006), "Kripke's Revenge", *Philosophical Studies* 128, pp. 669-682.

Burge, T. (1979), "Individualism and the Mental", *Midwest Studies in Philosophy* 4 (1), pp. 73-122.

Burge, T. (1986), "Individualism and Psychology," *Philosophical Review* 95, pp. 3-45.

Burge, T. (2010), *Origins of Objectivity*, Oxford University Press.

Carroll, J. W. (2010), "Laws of Nature", *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2012/entries/laws-of-nature/>>.

Chalmers, D. J. (2002), "The Components of Content", in D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press.

Chalmers, D. J. (2006), "Two-Dimensional Semantics", in E. Lepore & B. Smith (eds.), *The Oxford Handbook of Philosophy of Language*, Oxford University Press.

Clapp, L. (1995), "How to be Direct and Innocent: A Criticism of Crimmins and Perry's Theory of Attitude Ascriptions", *Linguistics and Philosophy* 18, pp. 529-565.

Crimmins, M. (1992), *Talk About Beliefs*, Cambridge, MA: MIT Press.

Crimmins, M. and Perry, J. (1989), "The Prince and the Phone Booth: Reporting Puzzling Beliefs", *Journal of Philosophy* 86 (12), pp. 685 - 711.

DePaul, M. & Ramsey, W. (eds.) (1998), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman & Littlefield.

Devitt, M. (1996), *Coming to Our Senses: A Naturalistic Program for Semantic Localism*, Cambridge: Cambridge University Press.

Dilworth, J. (2003), "A Refutation of Goodman's Type-Token Theory of Notation", *Dialectica* 57 (3), pp. 330-336.

Edwards, K. (2009), "What Concepts Do", *Synthese* 170, pp. 289-310.

Edwards, K. (2010), "Concept Referentialism and the Role of Empty Concepts", *Mind and Language* 25 (1), pp. 89-118.

Evans, G. (1982), *The Varieties Of Reference*, Oxford University Press.

Fodor J. A. (1994), *The Elm and the Expert*, Cambridge, MA: MIT Press.

Fodor, J. A. (1975), *The Language Of Thought*, Harvard University Press.

Fodor, J. A. (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.

Fodor, J. A. (1990), *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.

Fodor, J. A. (1998), *Concepts: Where Cognitive Science Went Wrong*, Oxford: Oxford University Press.

Fodor, J. A. (2004), "Having Concepts: A Brief Refutation of the Twentieth Century", *Mind and Language* 19 (1), pp. 29-47.

Fodor, J. A. (2008), *LOT 2: The Language of Thought Revisited*, New York: Oxford University Press.

Fodor, J. A. and Lepore, E. (1992), *Holism: A Shopper's Guide*, Cambridge, MA: Basil Blackwell.

Fodor, J. A. and Lepore, E. (1993), "Reply to Block and Boghossian", *Mind and Language* 8 (1), pp. 41-48.

Fodor, J. A. and Lepore, E. (1996), "The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes", *Cognition* 58, pp. 253–270.

Fodor, J. A. and Lepore, E. (2002), *The Compositionality Papers*, Oxford University Press.

Forbes, G. (1990), "The Indispensability of Sinn", *Philosophical Review* 99, pp. 535–563.

Forbes, G. (1993), "Reply to Marks", *Philosophical Studies* 69, pp. 281–295.

Frege, G. (1918-19/1956), "The Thought", *Mind* 65, pp. 287-311.

Glock, H. J. (2009), "Concepts: Where Subjectivism Goes Wrong", *Philosophy* 84 (1), pp. 5-29.

Goldman, A. (2007), "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status", *Grazer Philosophische Studien* 74, pp. 1-26.

Goldman, A. and Pust, J. (1998), "Philosophical Theory and Intuitive Evidence", in M. DePaul & W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman & Littlefield, pp. 201–239.

Goodman, N. (1968), *Languages of Art*, Indianapolis: Bobbs-Merrill Co.

Harman, G. (1993), "Meaning Holism Defended", in J. A. Fodor & E. Lepore (eds.), *Holism: A Consumer Update*, Amsterdam: Rodopi.

Harman, G. (1999), *Reasoning, Meaning, and Mind*, Oxford University Press.

Hempel, C. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.

Hempel, C. and Oppenheim, P. (1948), "Studies in the Logic of Explanation", *Philosophy of Science* 15, pp. 135-175.

Jackson, F. (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford University Press.

Jackson, F. (2007a), “Reference and Description From the Descriptivists' Corner” *Philosophical Books* 48 (1), pp. 17-26.

Jackson, F. (2007b), “On Not Forgetting the Epistemology of Names”, *Grazer Philosophische Studien* 74 (1), pp. 239-250.

Kaplan, D. (1989), “Demonstratives”, in J. Almog et al. (eds.), *Themes from Kaplan*, Oxford University Press.

Keil, F. (1989), *Concepts, Kinds and Cognitive Development*, Cambridge, MA: MIT Press.

Kripke, S. (1972/1980), *Naming and Necessity*, Harvard University Press.

Kripke, S. (1979), “A Puzzle about Belief”, in A. Margalit (ed.), *Meaning and Use*, Dordrecht: Reidel, pp. 239-283.

Kuhn, T. (1962), *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Laurence, S. and Margolis, E. (1999), “Concepts and Cognitive Science”, in E. Margolis & S. Laurence (eds.), *Concepts: Core Readings*, MIT Press.

Laurence, S. and Margolis, E. (2002), “Radical Concept Nativism”, *Cognition* 86, pp. 25-55.

Laurence, S. and Margolis, E. (2003), “Concepts and Conceptual Analysis”, *Philosophy and Phenomenological Research* 67, pp. 253-282.

Laurence, S. and Margolis, E. (2007), “The Ontology of Concepts – Abstract Objects or Mental Representations?”, *Noûs* 41 (4), pp. 561-593.

Laurence, S. and Margolis, E. (2011), “Concepts”, *The Stanford Encyclopedia of Philosophy*, URL = <<http://plato.stanford.edu/archives/fall2011/entries/concepts/>>.

Machery, E. (2009), *Doing Without Concepts*, New York: Oxford University Press.

Malt, B. (1989), "An On-line Investigation of Prototype and Exemplar Strategies in Classification", *Journal of Experimental Psychology: Learning, Memory and Cognition* 15, pp. 539-555.

Margolis, E. (1998), "Implicit Conceptions and the Phenomenon of Abandoned Principles", *Philosophical Issues* 9, pp. 105-114.

Margolis, E. (1999), "How to Acquire a Concept", *Mind and Language* 13, pp. 347-369.

Millikan, R. (1984), *Language, Thought and Other Biological Categories*, Cambridge: MIT Press.

Millikan, R. (1993), *White Queen Psychology and Other Essays for Alice*, Cambridge: MIT Press.

Millikan, R. (1998), "A Common Structure for Concepts of Individuals, Stuffs, and Kinds: More Mama, More Milk, and More Mouse", *Behavioral and Brain Sciences* 21 (1), pp. 55-65.

Millikan, R. (2000), *On Clear and Confused Ideas*, Cambridge: Cambridge University Press.

Millikan, R. (2011), "Loosing the Word-Concept Tie", *Aristotelian Society Supplementary Volume* 85 (1), pp. 125-143.

Murphy, G. L. (2002), *The Big Book of Concepts*, Cambridge, MA: MIT Press.

Osherson, D. N. and Smith, E. (1981), "On the Adequacy of Prototype Theory as a Theory of Concepts", *Cognition* 9, pp. 35-58.

Peacocke, C. (1992), *A Study of Concepts*, MIT Press.

Peacocke, C. (1998a), "Implicit Conceptions, Understanding and Rationality", *Philosophical Issues* 9, pp. 43-88.

Peacocke, C. (1998b), "Implicit Conceptions, the 'A Priori', and the Identity of Concepts", *Philosophical Issues* 9, pp. 121-148.

Peacocke, C. (2005), "Rationale and Maxims in the Study of Concepts", *Noûs* 39 (1), pp. 167-178.

- Peacocke, C. (2007), *Truly Understood*, Oxford: Oxford University Press.
- Pereboom, D. (1995), “Conceptual Structure and the Individuation of Content”, *Philosophical Perspectives* 9, pp. 401-428.
- Perry, J. (2000), *The Problem of the Essential Indexical and Other Essays*, 2nd ed., Stanford: CSLI Publications.
- Piccinini, G. and Scott, S. (2006), “Splitting Concepts”, *Philosophy of Science* 73, pp. 390-409.
- Pietroski, P. (1992), “Intentional and Teleological Error”, *Pacific Philosophical Quarterly* 73, pp. 267–281.
- Prinz, J. (2002), *Furnishing the Mind*, Cambridge, MA: MIT Press.
- Prinz, J. and Clark, A. (2004), “Putting Concepts to Work: Some Thoughts for the Twenty First Century”, *Mind and Language* 19 (1), pp. 57-69.
- Quine, W. V. O. (1951/1953), “Main Trends in Recent Philosophy: Two Dogmas of Empiricism”, *Philosophical Review* 60 (1), pp. 20-43.
- Recanati, F. (1993), *Direct Reference: From Language to Thought*, Oxford: Blackwell.
- Recanati, F. (2009), “De Re and De Se”, *Dialectica* 63 (3), pp. 249-269.
- Rey, G. (1983), “Concepts and Stereotypes”, *Cognition* 15, pp. 237-262.
- Rey, G. (1985), “Concepts and Conceptions: a Reply to Smith, Medin and Rips”, *Cognition* 19, pp. 297-303.
- Rey, G. (1994), “Concepts”, in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell.
- Rey, G. (2004), “Fodor’s Ingratitude and Change of Heart?”, *Mind and Language* 19 (1), pp. 70-84.
- Rey, G. (2005), “Philosophical Analysis as Cognitive Psychology: the Case of Empty Concepts”, in H. Cohen & C. Lefebvre (eds.), *Handbook of Categorization in Cognitive Science*, Dordrech: Elsevier, pp. 71-89.

Rey, G. (2009a), “Review of Edouard Machery, *Doing Without Concepts*”, *Notre Dame Philosophical Reviews* 7.

Rey, G. (2009b), “Concepts, Defaults and Internal Asymmetric Dependencies: Distillations of Fodor and Horwich”, in N. Kompa, C. Nimtz & C. Suhm (eds.), *The A Priori and Its Role in Philosophy*, Paderborn: Mentis, pp. 185-204.

Richard, M. (1990), *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*, Cambridge: Cambridge University Press.

Richard, M. (1997), “What Does Commonsense Psychology Tell Us About Meaning?” *Noûs* 31, pp. 87-114.

Rips, L. (1989), “Similarity, Typicality and Categorization”, in S. Vosniadou & A. Ortony (eds.), *Similarity and Analogical Reasoning*, Cambridge: Cambridge University Press, pp. 21-60.

Rupert, R. (2008), “Frege’s Puzzle and Frege Cases: Defending a Quasi-Syntactic Solution”, *Cognitive Systems Research* 9, pp. 76-91.

Sainsbury, R. M. (2010), “Paderewski Variations”, *Dialectica* 64 (4), pp. 483-550.

Sainsbury, R. M. and Tye, M. (2011), “An Originalist Theory Of Concepts”, *Proceedings of the Aristotelian Society Supplementary Volume* LXXXV, pp. 101-124.

Sainsbury, R. M. and Tye, M. (2012), *Seven Puzzles of Thought and How to Solve Them: an Originalist Theory of Concepts*, Oxford University Press.

Salmon, N. (1986), *Frege's Puzzle*, Cambridge, MA: MIT Press.

Salmon, N. (1989), “Illogical belief”, *Philosophical Perspectives* 3, pp. 243-285.

Salmon, N. (2006), “The resilience of illogical belief”, *Noûs* 40, pp. 369-375.

Saul, J. (1998), “The Pragmatics of Attitude Ascription”, *Philosophical Studies* 92 (3), pp. 363-389.

Saul, J. (1999a), “The Road to Hell: Intentions and Propositional Attitude Ascription”, *Mind and Language* 14 (3), pp. 356–375.

Saul, J. (1999b), “The Best of Intentions: Ignorance, Idiosyncrasy, and Belief Reporting”, *Canadian Journal of Philosophy* 29 (1), pp. 29 - 47.

Schiffer, S. (1987), “The ‘Fido’-Fido theory of belief”, *Philosophical Perspectives* 1, pp. 455-480.

Schiffer, S. (1990), “The Mode-of-Presentation Problem”, in C. Anderson & J. Owens (eds.), *Propositional Attitudes*, Stanford: CSLI.

Schiffer, S. (1992), “Belief Ascription”, *Journal of Philosophy* 89, pp. 499-521.

Schiffer, S. (1996), “The Hidden-Indexical Theory's Logical-Form Problem: A Rejoinder”, *Analysis* 56, pp. 92-97.

Schiffer, S. (2006), “A Problem for a Direct-Reference Theory of Belief Reports”, *Noûs* 40, pp. 361-368.

Schneider, S. (2005), “Direct Reference, Psychological Explanation, and Frege Cases”, *Mind and Language* 20, pp. 423-447.

Schneider, S. (2009a), “The Nature of Symbols in the Language of Thought”, *Mind and Language* 24, pp. 523-553.

Schneider, S. (2009b), “LOT, CTM, and the Elephant in the Room” *Synthese* 170 (2), pp. 235-250.

Schneider, S. (2011), *The Language Of Thought: A New Philosophical Direction*, Harvard: MIT Press.

Schroeter, L. (2004), “The Limits of Conceptual Analysis”, *Pacific Philosophical Quarterly* 85 (4), pp. 425-453.

Sider, T. (1995), “Three Problems for Richard's Theory of Belief Ascriptions”, *Canadian Journal of Philosophy* 25, pp. 487-513.

Smith, J. D. and Minda, J. P. (1998), “Prototypes in the Mist: The Early Epochs of Category Learning”, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, pp. 1411-1430.

Soames, S. (1987), “Substitutivity”, in J. Thomson (ed.), *On Being and Saying*, Cambridge: MIT Press, pp. 99-132.

Soames, S. (1995), “Beyond Singular Propositions”, *Canadian Journal of Philosophy* 25, pp. 515-549.

- Soames, S. (2002), *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, New York: Oxford University Press.
- Soames, S. (2005), *Reference and Description*, Princeton: Princeton University Press.
- Soames, S. (2006a), “Précis of Beyond Rigidity”, *Philosophical Studies* 128 (3), pp. 645 - 654.
- Soames, S. (2006b), “Replies to Critics”, *Philosophical Studies* 128 (3), pp. 711-738.
- Spelke, E. S. (1990), “Principles of Object Segregation”, *Cognitive Science* 14, pp. 29-56.
- Stalnaker, B. (1984), *Inquiry*, Cambridge University Press.
- Stalnaker, B. (2008), *Our Knowledge of the Internal World*, Oxford University Press.
- Weatherson, B. (2003), “What Good Are Counterexamples?”, *Philosophical Studies* 115, pp. 1-31.
- Weiskopf, D. (2009a), “Atomism, Pluralism, and Conceptual Content”, *Philosophy and Phenomenological Research* 79, pp. 130–162.
- Weiskopf, D. (2009b), “The Plurality of Concepts”, *Synthese* 169 (1), pp. 145–173.
- Weiskopf, D. and Bechtel, W. (2004), “Remarks on Fodor on Having Concepts”, *Mind and Language* 19 (1), pp. 18-56.
- Wetzel, L. (2006), “Types and Tokens”, *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2011/entries/types-tokens/>>.
- Williams, D. C. (1953), “The Elements of Being”, *Review of Metaphysics* 7, pp. 3–18, 171-192.
- Williamson, T. (2007), *The Philosophy of Philosophy*, Oxford University Press.
- Woodward, J. (2010), “Scientific Explanation”, *The Stanford Encyclopedia of Philosophy*, URL = <<http://plato.stanford.edu/archives/spr2010/entries/scientific-explanation/>>.

